Affiner la description des termes dans les dictionnaires généraux : l'apport d'un corpus de vulgarisation

Amélie Josselin-Leray[1]

Résumé

L'article vise à démontrer dans quelle mesure le recours à un corpus de vulgarisation peut être utile pour améliorer le recensement et la description des unités terminologiques dans les dictionnaires de langue générale. Après avoir montré quelle place les corpus occupent en lexicographie générale et comment ils sont utilisés dans le processus de la rédaction dictionnairique, l'article détaille la méthodologie de l'étude, qui porte sur le domaine de la volcanologie. L'analyse confronte les données de six dictionnaires généraux (deux dictionnaires unilingues anglais, deux dictionnaires unilingues français, et deux dictionnaires bilingues anglais/français) à celles trouvées dans deux corpus de vulgarisation (un corpus comparable anglais/français, et un corpus traduit bi-directionnel). La première partie de l'analyse s'intéresse à la macrostructure des dictionnaires, en établissant combien de termes parmi 220 choisis dans le corpus comparable (110 par langue) sont recensés et de quel type de termes il s'agit. La deuxième partie de l'analyse se concentre sur la microstructure des dictionnaires, en montrant comment l'utilisation du corpus peut permettre d'améliorer les indications grammaticales, les définitions et les équivalents d'un certain nombre de termes.

Mots-clés : terminologie – lexicographie générale – corpus comparable – corpus traduit bi-directionnel – vulgarisation – volcanologie

***

Abstract

The aim of the paper is to show how useful a popular-science corpus can be to improve the inventory and description of terminological units in general-purpose dictionaries. The article first discusses the role of corpora in general-purpose lexicography and explains how they are used in the dictionary-writing process. In the second section of the article, the methodology used in this study is described. The study, which focuses on the field of volcanology, compares the data found in six general-purpose dictionaries (two English monolingual dictionaries, two French monolingual dictionaries, and two French/English bilingual dictionaries) to those found in two popular-science corpora (one French/English comparable corpus, and one bi-directional translation corpus). The first part of the analysis deals with the macrostructure of the dictionaries, and determines how many terms (among the 220 picked in the comparable corpus–110 per language) are listed and what types of terms they are. The second part focuses on the microstructure, and shows how resorting to a corpus can improve the grammatical indications, definitions and equivalents of a number of terms.

Key-words: terminology – general-purpose lexicography – comparable corpus – bi-directional translation corpus – popular science – volcanology

[1] Université Toulouse 2 – le Mirail, CLLE-ERSS, UMR 5263 : josselin@univ-tlse2.fr

Introduction

Lexicographie oblige, quelques précisions lexicales s'imposent à l'orée de ce travail. Nous entendons par *terme* une unité lexicale relevant d'une langue de spécialité. Plus précisément, nous faisons nôtre la définition de L'Homme [2002a : 20] :

> **terme** désigne toute unité lexicale dont le sens peut être appréhendé et décrit en fonction des rapports de cette unité avec un domaine de la connaissance humaine. Les termes sont des noms (ex. : *commerce de détail, logiciel*), des verbes (ex. *configurer, enchérir*), des adjectifs (ex. : *compatible, publiposté*) ou des adverbes (ex. : *linguistiquement*). Les termes peuvent être simples (ex. : *hypermarché, souris*) ou complexes (ex. : *moteur à explosion, système d'exploitation*). Les termes complexes sont généralement décrits comme des termes simples même si leur sens est compositionnel.

Cette définition a pour intérêt qu'elle s'articule autour de trois dimensions essentielles du terme, à savoir : concept, domaine et morphologie. Elle en oublie peut-être une, qui est celle du fonctionnement en discours. En effet, comme n'importe quelle unité lexicale, le terme s'insère dans le discours. La dimension discursive du terme est d'ailleurs manifeste à travers le mot qu'utilisent Boulanger & L'Homme [1991] pour désigner les terminologies, à savoir le mot *technolecte*.

Si nombre d'auteurs (comme Mortureux [1995 : 22-23]) s'accordent à dire que l'on peut définir le *terme* comme l'unité lexicale utilisée par les spécialistes lorsqu'ils **parlent** de leur domaine, il faut reconnaître que les termes ne sont pas cantonnés aux seuls discours spécialisés (ou communication entre experts) : ils circulent, comme le fait très justement remarquer ce même auteur [*ibid.* : 22]. Et c'est justement parce que les termes circulent qu'ils se retrouvent dans les dictionnaires dits de langue générale. Même s'il s'agit là d'une sorte de paradoxe qu'il convient de souligner, force est de constater, d'une part, que les termes occupent une place non négligeable dans les dictionnaires de langue, place qui semble par ailleurs en constante expansion (cf. Mazière [1981 : 82] et Landau [1989 : 21]), et d'autre part, que personne ne semble remettre en cause cet état de fait : ni les lexicographes qui s'en enorgueillissent sur les quatrièmes de couverture de leurs ouvrages, ni les métalexicographes, comme Boulanger [2001 : 247], qui trouvent cela bien naturel, ni les utilisateurs qui, on le sait bien, veulent des dictionnaires qui contiennent les mots difficiles, donc les termes — c'est d'ailleurs ce que confirment les résultats d'un sondage effectué en 2003-2004[2], selon lesquels jusqu'à 54% des personnes interrogées ne se disent pas prêtes à acquérir un dictionnaire de langue générale qui ne contiendrait pas de mots scientifiques et techniques (Josselin-Leray [2005 : 330]).

Malgré tout, selon cette même enquête, les utilisateurs se disent dans l'ensemble plutôt insatisfaits du traitement réservé à ce type d'unité lexicale dans les dictionnaires, que ce soit au niveau de la macrostructure[3] ou de la microstructure [*ibid.* : 358-362]. Tandis que beaucoup d'études s'attachent à démontrer, de manière globale, comment le recours à un corpus électronique peut améliorer le traitement de la langue générale dans les dictionnaires généraux (Sinclair *et al*. [1987], Atkins [1994], Teubert [2001], Béjoint [2007]), il en est peu

[2] Sondage effectué auprès de 346 utilisateurs de dictionnaires unilingues (français ou anglais) et 301 utilisateurs de dictionnaires bilingues (anglais / français). Pour plus de détails sur cette enquête, voir Josselin-Leray [2005 : 249-363].

[3] La critique est facile et récurrente ; en effet, le dictionnaire est toujours jugé sur l'absence du mot ou du terme dont on a besoin.

qui viennent s'attacher au traitement de la langue spécialisée dans ces dictionnaires. Nous espérons donc ici combler un manque.

Après un bref exposé sur le rôle du corpus en lexicographie et en terminologie, nous détaillons la méthodologie de notre étude qui a consisté en l'analyse approfondie du traitement de la terminologie de la volcanologie dans six dictionnaires (deux dictionnaires unilingues en français, deux dictionnaires unilingues en anglais, deux dictionnaires bilingues anglais / français) à la lumière des données fournies par un corpus de vulgarisation de ce domaine. Dans une troisième partie, nous livrons les résultats de cette analyse concernant la macrostructure, puis dans une quatrième partie les résultats propres à la microstructure.

1. Dictionnaires, termes et corpus

1.1. Place du corpus en lexicographie générale

1.1.1. Lexicographie et corpus : cadre général

La démarche lexicographique du XXI[e] siècle est-elle fondamentalement différente de celle que décrivait John Sinclair en 1985 ? Selon lui, le lexicographe qui compile ses données s'appuie sur les trois sources suivantes : (1) son intuition de linguiste, (2) les dictionnaires préexistants, (3) des textes représentant la langue telle qu'elle est réellement utilisée. C'est toujours le cas aujourd'hui, à ceci près qu'avec l'avènement de l'informatique, la troisième source, désormais sous forme de corpus électronique, s'est taillé la part du lion et a pris le pas sur les deux premières (autrefois, l'introspection du lexicographe concernant le choix des unités et des sens à retenir primait sur la fréquence d'usage, comme l'explique Béjoint [2007 : 36]).

1.1.2. Termes et dictionnaires généraux : une prépondérance accrue du corpus

Comme nous l'avons déjà signalé, plusieurs études ont démontré l'intérêt que présentait le recours à un corpus pour la description lexicographique de la langue générale (notamment Atkins [1994] ou Sinclair [1987]). Ce recours semble plus crucial encore dans le cas des langues de spécialité, ou du moins, dans le cas fréquent du lexicographe ayant à traiter des langues de spécialité ou terminologies dans le cadre d'un dictionnaire général. En effet, comme l'expliquent Meyer & Mackintosh [1996 : 265], les terminologues peinent à procéder par introspection et sont donc très dépendants de leur documentation. Que dire alors du lexicographe « généraliste » qui travaille sur diverses langues de spécialité pour lesquelles il est pour la plupart du temps novice ? De plus, la tendance générale est au rapprochement des méthodes entre lexicographes et terminologues[4] étant donné l'évolution théorique majeure qu'a connue la terminologie au cours des quinze dernières années. En effet, ce qui a longtemps différencié la lexicographie de la terminologie était le sens de sa démarche : tandis que les lexicographes procédaient de manière sémasiologique (de la désignation au concept), les terminologues disaient procéder de manière onomasiologique (du concept à la désignation) ; depuis quelques années s'est développé le courant de la « terminologie textuelle » pour lequel « l'activité de construction d'une terminologie est désormais essentiellement une tâche d'analyse de corpus textuels », comme l'expliquent Bourigault & Slodzian [1999 : 30].

[4] Sur la question de la compatibilité des modèles et des méthodes entre lexicographie et terminologie, voir l'ouvrage collectif de L'Homme et Vandaele [2007], et en particulier le chapitre 1 : « Nouvelle lexicographie et nouvelles terminologies : convergences et divergences » (Béjoint [2007]).

1.2. Utilisation des corpus dans le processus de la rédaction dictionnairique

1.2.1. Quels corpus ?

Nous avons indiqué plus haut quels dictionnaires avaient innové en se servant de corpus électroniques pour la première fois. Aujourd'hui, qu'en est-il pour les dictionnaires usuels de langue générale ? Les dictionnaires sont parfois peu bavards sur les corpus utilisés lors de leur élaboration, et dans certains cas il est assez difficile d'établir s'ils reposent sur de réels corpus électroniques ou sur de simples collections de texte. De manière générale, on note une grande disparité entre les dictionnaires de langue française et les dictionnaires de langue anglaise. Béjoint [*ibid.* : 34] indique ainsi qu'il ne dispose d'aucune information sur les corpus des dictionnaires français les plus vendus, le *Petit Larousse* et le *Petit Robert*, « alors que les corpus utilisés par les dictionnaires [britanniques] sont connus, décrits et souvent même accessibles, en tout ou en partie ».

Lorsqu'on s'intéresse plus particulièrement aux corpus utilisés pour décrire les termes dans les dictionnaires généraux, le problème semble devenir encore plus épineux. Ainsi, comme le font remarquer Boulanger & L'Homme [1991 : 29] :

> La majorité des dictionnaires ne précise pas vraiment quelle est la provenance des contingents technolectaux.

Cette affirmation est toutefois nuancée, et il est vrai que certains dictionnaires ou lexicographes sont plus loquaces que d'autres sur les sources d'où sont tirés les termes inclus dans les dictionnaires. Par exemple, Candel [1979 : 114] commente en ces termes la richesse de la documentation du *TLF* :

> Le très grand nombre d'œuvres dépouillées par le *TLF* permet aux articles de ce dictionnaire d'être particulièrement développés, et de présenter un choix important d'exemples. Ainsi, alors que l'emploi de baril en métrologie n'est pas attesté par le [*Grand Robert*], ni par le [*Grand Larousse*], il est défini, dans le *TLF*, par un dictionnaire récent, [le] *Dictionnaire de l'économie contemporaine*, 1968, et illustré par un exemple extrait de *Pétroles naturels et artificiels*, de [...] 1931, une seconde définition étant extraite du complément au *dictionnaire de l'Académie* de 1842.

Ces deux exemples, ainsi que bien d'autres (voir notamment Boulanger & L'Homme [1991 : 29], à propos de la constitution des sources du *Grand Larousse*), montrent que, pour la sélection des termes à inclure dans le dictionnaire général, les dictionnaires font appel à deux sources : d'une part, à des *locuteurs* de niveaux de spécialisation différents, à savoir des lexicographes non spécialisés, des traducteurs spécialisés, des spécialistes du domaine (comme en témoignent d'ailleurs les listes des contributeurs recensés dans les parties introductives des dictionnaires) ; d'autre part, à des *textes* de nature différente et de degré de spécialisation différents.

1.2.2. Intégration du corpus aux diverses étapes de la rédaction dictionnairique

A quel(s) niveau(x) du processus de l'élaboration d'un dictionnaire le recours au corpus intervient-il ? L'émergence des corpus sous forme électronique a, semble-t-il, bouleversé les deux premières étapes de sélection des mots-vedettes et de recueil des données sur la langue

en usage (Roberts [1992 : 100], Svensén [1993 : 250-271)]. Peu de lexicographes se sont exprimés en détail sur la chronologie du processus. Seuls deux exemples sont assez parlants sur le sujet et nous semblent intéressants à évoquer même s'ils représentent des cas particuliers : un exemple concernant la lexicographie unilingue anglaise, et un exemple concernant la lexicographie bilingue (anglais / français).

Pour ce qui est des dictionnaires unilingues, Sinclair et ses collaborateurs [1987] ont décrit par le menu leur utilisation du corpus *The Birmingham Collection of Texts* pour divers aspects du dictionnaire pour apprenants *The Collins Cobuild*. Ainsi, d'après les informations données par Ramesh Krishnamurthy ("Senior Compiler") dans le chapitre intitulé "The Process of Compilation" [Sinclair, 1987 : 62-85]**,** il semble que le corpus soit utilisé dès la première étape décrite plus haut. En effet, Krishnamurthy explique qu'à quelques exceptions près, un mot-vedette figurant dans la version préliminairement dressée de la nomenclature mais dont aucune occurrence n'est par la suite trouvée dans le corpus principal n'est pas retenu pour figurer dans la nomenclature finale du dictionnaire. La politique générale est que le corpus est primordial, et que les ouvrages de référence existants ne sont que secondaires et consultés uniquement après l'analyse approfondie des données du corpus.

En ce qui concerne les dictionnaires bilingues, Atkins [1994] nous livre dans l'introduction même du dictionnaire *Oxford-Hachette* le mode d'emploi du corpus dans une partie intitulée « How do we use a corpus? ». Elle y explique que la première étape se fait de manière traditionnelle : dans chaque langue, des lexicographes établissent des sortes de pré-entrées dictionnairiques ou grilles à partir d'ouvrages de référence existants ; ces grilles sont ensuite traduites puis redonnées à d'autres lexicographes qui procèdent à des ajouts ou à des vérifications en se basant sur les données du corpus.

Après ce tour d'horizon de la place qu'occupe, dans tous les sens du terme, le corpus dans la lexicographie générale, il est temps d'expliciter les conséquences que nous en tirons pour le protocole à suivre pour notre propre étude.

2. Méthodologie de l'étude

2.1. Présentation du protocole

Pour l'ensemble de notre démarche, nous avons tenté de nous rapprocher du rôle du lexicographe du XXI^{e} siècle qui vit avec son temps, intégrant les nouvelles technologies d'une part et les changements théoriques affectant la terminologie d'autre part.

2.1.1. Choix du domaine, de la philosophie de la démarche et du type de corpus

Le but de notre étude étant de démontrer comment l'utilisation d'un corpus peut permettre d'affiner le recensement et la description d'un certain type d'unité lexicale dans les dictionnaires de langue générale, il fallait tout d'abord choisir un domaine d'application dont la terminologie soit d'une part relativement limitée et d'autre part assez populaire pour avoir droit de cité dans les dictionnaires généraux, qui reflètent la culture de la société qui les produit. C'est une des raisons pour lesquelles notre choix s'est porté sur le domaine de la volcanologie. La volcanologie est en effet un domaine tout particulièrement bien vulgarisé en France, et fait donc l'objet, semble-t-il, d'une attention privilégiée dans les dictionnaires français (par exemple, le volcanologue J.-M. Bardintzeff est désormais conseiller scientifique du *Petit Larousse*), ce qui corrobore le constat fait par Laurian [2006 : 357] :

> Le dictionnaire [bilingue] de petit format s'adressant à un vaste public doit représenter ou transporter une image de la science également grand public […] De domaines privilégiés en terrains de prédilection, on aura probablement un choix de disciplines qui correspondra aux attentes d'un public friand de connaissances limitées.

Il fallait ensuite établir l'orientation générale de notre démarche. En effet, plusieurs études métalexicographiques ont déjà tenté d'évaluer le traitement des termes dans les dictionnaires de langue générale. Une comparaison fine de toutes les méthodes utilisées aboutit à la conclusion suivante[5] : jusqu'à présent, trois méthodes différentes ont été utilisées par les chercheurs. La première consiste à utiliser les dictionnaires de langue générale comme point de départ, méthode qui se décline sous trois formes : (*i*) Wesemael & Wesemaël [2003] ont décidé d'analyser, dans la première partie de leur article, des néologismes scientifiques (issus de domaines scientifiques différents) cités par les auteurs du *Nouveau Petit Robert* de 2002 dans la préface ; (*ii*) d'autres auteurs, plus nombreux, procèdent en « prélevant » des échantillons dans telle ou telle lettre du dictionnaire, repérés par le biais d'une étiquette de domaine (Mazière [1981], Boulanger [2001]) ; (*iii*) enfin, un auteur (Rodriguez Reina [1999]) a dépouillé linéairement le dictionnaire général pour y trouver les termes d'un domaine précis (les termes maritimes). La deuxième méthode utilise comme point de départ des dictionnaires spécialisés du domaine. Plusieurs auteurs, dont Candel & Humbley [1997] et Thoiron [1998], relèvent des termes (dans les domaines respectifs de la géomorphologie pour les premiers et de la médecine pour le deuxième) dans des répertoires spécialisés, puis confrontent les données de ces ouvrages à celles tirées des dictionnaires généraux. La troisième méthode, utilisée par seulement deux études, se fonde sur des textes où circulent des mots du domaine, soit des articles de journal, soit des ouvrages de vulgarisation. C'est ainsi que Mazière [1981] étudie les termes d'un article du *Monde des Sciences* et que Wesemael & Wesemaël [2003] décortiquent un article de *Pour la Science*. Ces derniers ont également « parcouru […] le célèbre ouvrage de vulgarisation astronomique *Patience dans l'azur* (Reeves 1981) » [ibid. : 151].

On voit donc que les études faisant primer le contexte d'utilisation des termes, c'est-à-dire s'appuyant sur une forme de corpus, sont peu nombreuses ; et aucune étude n'utilise de corpus *électronique* pour l'analyse. Baser notre travail sur un corpus présente donc un caractère novateur, et s'inscrit dans la droite ligne des pratiques des lexicographes telles que nous les avons décrites plus haut. Nous avons jugé que le corpus adéquat pour la description de la terminologie de la volcanologie à des fins de lexicographie générale serait un corpus de vulgarisation. A l'instar de Delavigne [2001], il nous paraît en effet légitime d'utiliser des textes de vulgarisation comme support terminologique. Par ailleurs, Candel [1994 : 191] et Mortureux [1994 : 66] notent toutes deux la pertinence du recours à de tels textes dans le contexte de la lexicographie générale, expliquant que les lexicographes peuvent mettre immédiatement à profit les définitions et reformulations qui s'y trouvent, et que les exemples y figurant sont sans doute « plus éclairants » [Mortureux, *ibid*. : 72] que ceux trouvés dans les discours spécialisés.

2.1.2. L'utilisation de corpus de vulgarisation : la démarche pas à pas

Notre démarche s'est apparentée à ses débuts à celle de la terminologie textuelle. En effet, un corpus de vulgarisation a constitué le point de départ de notre analyse. Nous avons en réalité élaboré et utilisé deux types de corpus bilingues, dont nous présentons le contenu plus bas (section 2.2.1) : *(i)* un corpus comparable anglais-français, tel que le définissent

[5] Pour plus de détails sur cette comparaison, voir Josselin-Leray [2005 : 88-114].

Altenberg & Granger [2002 : 7-8] : « comparable corpora consist of original texts in each language, matched as far as possible in terms of text type, subject matter and communicative function », et *(ii)* un corpus traduit bi-directionnel anglais-français, tel que le définissent ces mêmes auteurs [ibid. : 8] :

> Translation corpora consist of original texts in one language and their translations into one or several other languages. If the translations go in one direction only (from language A to language B) they are unidirectional; if they go in both directions (from language A to language B and from language B to language A) they are bi-directional.

Grâce à l'analyseur syntaxique de corpus Syntex (cf. Bourigault *et al*. 2005), nous avons extrait du corpus comparable, pour chaque langue, une liste de candidats-termes (termes simples et termes complexes). Une analyse fine des données du corpus nous a permis d'établir une liste de termes dont nous avons ensuite étudié le traitement dans un certain nombre de dictionnaires généraux ; mais cette analyse a impliqué un retour fréquent au corpus afin de pouvoir bénéficier de l'éclairage des contextes. En effet, notre démarche n'est pas figée dans un seul sens, qui serait celui d'un mouvement du corpus vers les dictionnaires. Ce va-et-vient constant entre corpus et dictionnaires a caractérisé l'ensemble du travail. Nous avons synthétisé notre démarche, que l'on pourrait qualifier de circulaire, dans le schéma qui suit (fig.1). Précisons enfin que le corpus traduit n'a été utilisé que dans un second temps, afin de rechercher ou de vérifier les équivalents des termes identifiés dans le corpus comparable.

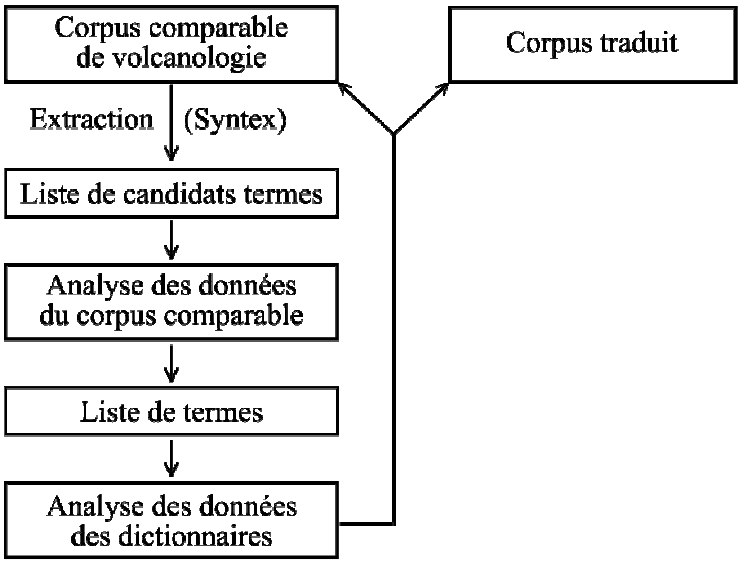

Figure 1. Schématisation de la démarche d'analyse

Prenons juste un exemple pour illustrer la complexité de la démarche, celui du verbe *erupt*. Ce terme a été retenu parmi les candidats-termes extraits par Syntex : il s'agissait du verbe le plus fréquent (nombre d'occurrences fourni par Syntex : 554). En dépouillant les données des dictionnaires bilingues étudiés (qui seront présentés en détail dans la section 2.2.2), tout en parcourant simultanément des données du corpus comparable, nous avons pu constater une disparité majeure : alors que les dictionnaires bilingues présentent ce verbe de manière fort simple, comme en attestent les entrées suivantes :

Erupt (…) vi
(a) **(of volcano) entrer en éruption (b) (of violence, anger, sound) éclater; (of person) exploser** the stadium erupted in a huge roar **le stade a éclaté d'un énorme rugissement (c) (of teeth) percer; (of spot) sortir** his face erupted in spots **son visage s'est couvert de boutons**

Figure 2. Entrée *erupt* du *Harrap's Shorter* sur CD-ROM (2000)

Erupt (…) intransitive verb
1 [volcano] entrer en éruption;
2 figurative [war, violence, gunfire] éclater; [laughter, cry] jaillir, éclater; [person] (with anger) éclater (with de);
3 medicine [rash] apparaître; dentistry [tooth] percer.

Figure 3. Entrée *erupt* du *Dictionnaire Oxford-Hachette* sur CD-ROM (1996)

Les données du corpus révèlent un fonctionnement du verbe bien plus complexe. En parcourant les contextes fournis par Syntex, nous sommes en effet tombée sur les phrases suivantes :

(1) In July 1986, just as it appeared that Puu Oo would start the 48th episode, **magma** began **to erupt** from new fissures at the base of the cone**.**
(2) During the subaerial shield building stage **aa and pahoehoe erupt** from the summit area and rift zones**.**
(3) After periods of volcanic quiescence lasting as long as hundreds of thousands to even a million years, some Hawaiian **volcanoes erupt** alkalic volcanic **products**.
(4) **Pillow lavas** are volumetrically the most abundant type because **they are erupted** at mid ocean ridges.

Les contextes (1) et (2) montrent que des sujets autres que *volcano*, à savoir *magma* ou encore des types de lave comme *aa and pahoehoe* peuvent être utilisés pour le verbe *erupt*. Or, il n'y a trace de ces sujets dans aucun des deux dictionnaires bilingues.

De plus, dans les contextes (3) et (4), le verbe *erupt* est utilisé de manière *transitive,* emploi qui n'est mentionné dans aucun des deux dictionnaires bilingues analysés, et qui n'est mentionné (de manière très sommaire) que dans le dictionnaire unilingue *American Heritage College Dictionary* (1996), de la manière suivante : « --tr. To force out violently ».

Enfin, dans ces quatre cas, l'équivalent français du verbe *erupt* n'est pas *entrer en éruption*. Etant donné que les quatre extraits proviennent de la même source[6], nous nous sommes demandée s'il ne s'agissait pas là d'un phénomène idiolectal. Nous avons donc analysé l'ensemble des 554 occurrences du verbe dans la partie anglaise du corpus comparable, ce qui nous a permis de constater que cette construction transitive n'était pas limitée à une seule source. Nous avons donc établi la proportion de chaque construction du verbe, avons cherché des équivalents potentiels dans la partie française du corpus comparable, puis avons analysé les 44 occurrences de ce verbe dans le corpus traduit, afin d'étudier les traductions possibles, le tout en regardant également comment ce verbe était présenté dans les dictionnaires unilingues anglais.

[6] *A Teacher's Guide to the Geology of Hawaii Volcanoes National Park*, dont le texte est consultable sur Internet à l'adresse suivante (au 31.08.2009) : http://volcano.oregonstate.edu/vwdocs/vwlessons/atg.html

Examinons à présent plus en détail notre manière de procéder pour l'analyse de la macrostructure et de la microstructure du dictionnaire.

2.1.3. Analyse de la macrostructure et de la microstructure

L'extraction automatique de termes effectuée par Syntex nous a livré une liste qui n'était pas exploitable directement car elle contenait beaucoup de « bruit », et un travail de sélection des termes potentiels s'est avéré nécessaire. Par exemple, Syntex nous livrait comme résultats « bruts » de l'extraction 7095 noms communs simples en anglais et 7836 en français[7]… En ayant établi de manière rigoureuse une liste précise de critères de rétention et de rejet, nous sommes parvenue, non sans mal, à un ensemble de 110 termes anglais et de 110 termes français, qui ne sont pas nécessairement des équivalents traductionnels les uns des autres[8]. Parmi ces 110 termes, nous avons retenu, pour chacune des langues : 52 noms simples ; 36 syntagmes nominaux, dont certains sont des noms composés, et d'autres des collocations ; 16 adjectifs ; 6 verbes, syntagmes verbaux, ou collocations verbales.

Ces chiffres (110, 52…) n'ont pas été choisis de manière arbitraire : afin de choisir combien de termes retenir pour l'analyse, nous avons décidé de tenter de comptabiliser combien de termes du domaine de la volcanologie étaient inclus en moyenne dans les dictionnaires généraux que nous étudions. Nous avons donc procédé à un dénombrement approximatif dans chacun des dictionnaires unilingues étudiés. Le décompte précis des termes dans les dictionnaires généraux n'est pas chose aisée, comme l'ont constaté bon nombre de nos prédécesseurs, notamment Thoiron [1998 : 624] qui, s'intéressant à la terminologie médicale, explique qu'« il ne suffit pas de compter les étiquettes *Médecine* de la version publiée pour connaître le nombre de termes médicaux dans [*OXHA*] ». La méthode de repérage des termes grâce à une étiquette de domaine étant de toute manière totalement invalide dans notre cas (puisque l'étiquette *Volcanologie* n'existe dans aucun des six dictionnaires que nous analysons), nous avons procédé par recherche de mots-clés en texte intégral (pour l'anglais : *volcano* et ses dérivés, *erupt* et ses dérivés, *lava* ; pour le français, *volcan* et ses dérivés, *éruption* et ses dérivés, ainsi que *lave*). A ce stade, il ne s'agissait pas d'effectuer un comptage exact et précis, mais d'avoir uniquement un ordre de grandeur de ce que proposent les dictionnaires dans le domaine de la volcanologie pour que notre propre liste de termes soit d'un ordre de grandeur comparable.

Pour l'analyse de la macrostructure des dictionnaires existants, les 110 termes français ont fait l'objet d'une recherche dans les dictionnaires unilingues français et d'une recherche dans les deux sections des dictionnaires bilingues anglais-français — ils ont été recherchés à la fois comme entrées en langue source dans la section français-anglais, et comme équivalents en langue cible dans la section anglais-français. Les 110 termes anglais ont été recherchés dans les dictionnaires unilingues anglais ainsi que dans les deux sections des dictionnaires bilingues anglais-français. Les dictionnaires analysés ont été en partie choisis parce qu'ils disposent d'une version CD-ROM assez élaborée, qui permet notamment la recherche en texte intégral. La plupart d'entre eux disposent de plus d'une fonction de recherche utilisant les opérateurs booléens tels que *et* ou *and*, ce qui permet de trouver plus facilement les unités terminologiques complexes (*volcan en éruption*, *(to) spew ash*, etc.). Cette méthode est certes avantageuse car elle permet l'accès à une information plus complète et plus détaillée (il est

[7] Ceci peut laisser à penser que la démarche dite d'apprentissage endogène utilisée par Syntex (c'est-à-dire une démarche ne reposant sur aucune ressource lexicale préalable) n'est pas très efficace pour extraire la terminologie d'un domaine donné.

[8] Le relevé des 110 termes, ainsi que la liste des critères de sélection et l'explication des divers problèmes afférents à la constitution et à son application ne peuvent être présentées de manière satisfaisante en peu de lignes. Nous renvoyons donc le lecteur à Josselin-Leray [2005 : 413-450].

possible d'aller dénicher tous les secrets que le dictionnaire ne livre pas de prime abord), mais elle a l'inconvénient d'être fort fastidieuse et il faut veiller à ne pas se perdre dans les méandres du dictionnaire : par exemple, la recherche en texte intégral de l'adjectif *volcanique* dans le *NPR* révèle que ce terme figure dans 56 entrées du dictionnaire... Un tri s'imposait donc, et certains critères devaient de nouveau être établis pour considérer que le terme était réellement répertorié dans le dictionnaire. Nous y reviendrons plus en détail. Nous analyserons tout d'abord le traitement des 110 termes anglais et des 110 termes français au niveau de la macrostructure, puis le traitement de certains de ces 110 termes pour des points particuliers de la microstructure.

2.2. Présentation des outils

2.2.1. Les corpus de vulgarisation : bref aperçu

Présentons maintenant à grands traits les caractéristiques générales des deux corpus constitués par nos propres soins. Le corpus comparable et le corpus traduit comportent des textes en langue anglaise (variétés britannique, américaine et canadienne) et en langue française (variétés française et canadienne) portant sur le domaine de la volcanologie. Le corpus comparable comporte environ 400 000 mots par langue, soit un total d'environ 800 000 mots tandis que le corpus traduit contient environ 100 000 mots par langue, soit 200 000 mots au total. Nos deux corpus sont donc d'une taille adaptée à nos besoins si l'on se réfère aux conseils de Bowker & Pearson [2002 : 48] : "In our experience, well-designed corpora that are anywhere from about ten thousand to several hundreds of thousands of words in size have proved to be exceptionally useful in LSP studies." La période couverte par les deux corpus est d'une vingtaine d'années (de la fin des années 1970 jusqu'aux années 2000). Les textes inclus sont des textes écrits, entiers, de source fiable et que l'on peut considérer comme des « textes de vulgarisation », même si l'appellation recouvre en réalité plusieurs types de discours (du discours du quotidien généraliste au discours de semi-vulgarisation, en passant par exemple par le discours pédagogique). Les tableaux qui suivent (tableaux 1 et 2) présentent de manière récapitulative les divers niveaux de subdivision du corpus selon le type de texte (nom des subdivisions et taille en nombre de mots) et mettent côte à côte divers exemples de documents anglais et français comparables ou traduits.

<table>
<tr><th colspan="3">Type de document</th><th>Ex. de titres français</th><th>Nombre de mots</th><th>Ex. de titres anglais</th><th>Nombre de mots</th></tr>
<tr><td rowspan="5">TEXTES</td><td rowspan="5">Hors Presse</td><td>Manuels académiques</td><td>Eléments de Géologie</td><td>12 689</td><td>Essentials of Geology</td><td>10 590</td></tr>
<tr><td>Ouvrages de vulgarisation</td><td>Les Volcans et leurs Secrets</td><td>81 151</td><td>General-Interest Publications of the USGS[9] on Volcanoes</td><td>91 208</td></tr>
<tr><td>Documents du Web</td><td>Le Volcanisme de Djibouti</td><td rowspan="2">26 732</td><td>How Volcanoes Work</td><td rowspan="2">21 603</td></tr>
<tr><td>Exposition</td><td>Des Volcans et des Hommes</td><td>∅</td></tr>
</table>

[9] United States Geological Survey

	Presse	Quotidiens généralistes	*Le Monde* *La Presse*	145 369	*The Guardian (GB)* *The New-York Times (US)* *The National Post (CD)*	148 642
		Magazines de vulgarisation	*Science et Vie* *Science et Avenir*	58 609	*New Scientist (GB)* *Science News (US)*	58 633
		Magazines de semi-vulgarisation	*Pour la Science*	15 689	*Scientific American*	16 094
		Magazines de voyage / découverte	*Géo* *Terre Sauvage*	56 446	*Earth* *National Geographic*	54 152
GLOSSAIRES				**3 227**		**4 686**
TOTAL				**400 486**		**405 546**

Tableau 1. Composition du corpus comparable

	Partie française				**Partie anglaise**		
Type	**Titre**	**Nombre de mots**	**Date**	**Sens de la traduction**	**Titre**	**Nombre de mots**	**Date**
Magazine de semi-vulgarisation	*Pour la Science*	70 132	1979 - 1999	anglais-français	*Scientific American*	67 762	1979 - 1999
Dépêche scientifique	*Volcanisme et Volcans d'Auvergne*	12 482	2002	français-anglais	*Volcanism and Volcanoes of Auvergne*	11 804	1999
Ouvrage de vulgarisation	*Les Feux de la Terre-Histoires de Volcans*	28 092	1991	français-anglais	*Volcanoes-Fire from the Earth*	26 268	1993
Total		**110 712**				**105 834**	

Tableau 2. Composition du corpus traduit

2.2.2. Les dictionnaires analysés

Le tableau qui suit présente de manière synthétique les six dictionnaires généraux analysés :

	Dictionnaires unilingues				**Dictionnaires bilingues**	
Titre	*Petit Larousse*	*Nouveau Petit Robert*	*American Heritage College Dictionary*	*New Oxford Dictionary of English*	*Dictionnaire Oxford-Hachette*	*Harrap's Shorter Dictionary*
Abréviation	*PL*	*NPR*	*AH*	*NODE*	*OXHA*	*HAR*
Date de publication	2002	2001	1996	2000	1996	2000
Support	CD-ROM	CD-ROM	CD-ROM	CD-ROM	CD-ROM	.CD-ROM
Variété de langue	FR*[10]	FR*	US*	GB*	FR-US-GB	FR-US-GB
Taille de la nomenclature[11]	60 000 mots et leurs 300 000 sens	59 000 mots et leurs sens. Les locutions et les expressions	over 200,000 definitions	350,000 words, phrases and definitions	350 000 mots et expressions	305 000 mots et expressions
Corpus électronique utilisé	Non renseigné	Non renseigné	Non renseigné	*British National Corpus* and other corpora	*The Bank of French*	Non renseigné

Tableau 3. Présentation des six dictionnaires généraux analysés

Ces dictionnaires ont été choisis sur la base d'un certain nombre de critères communs à tous les dictionnaires, ou seulement propres à certains d'entre eux.

• Critères communs à tous les dictionnaires :

– il s'agit de dictionnaires sur CD-ROM ; ce support a été choisi car il témoigne d'une plus grande modernité et permet a priori d'effectuer des recherches plus pointues qu'un dictionnaire papier ;

– il s'agit de dictionnaires de taille ou de nomenclature comparables : ce sont des dictionnaires que l'on peut qualifier d'usuels (ce qui correspond approximativement au format "college" pour les dictionnaires américains), autrement dit des dictionnaires dont l'équivalent papier est un gros volume ; précisons toutefois que les traditions et les politiques éditoriales étant propres à chaque pays, il est impossible d'effectuer un appariement parfait. Outre le fait que chaque dictionnaire a sa propre manière de comptabiliser et de présenter le nombre de mots qu'il contient (l'*AH* est notamment très flou à ce sujet, comme en témoigne la ligne « taille de la nomenclature » du tableau ci-dessus), il faut garder à l'esprit le fait que la présentation microstructurelle peut varier d'un dictionnaire à l'autre (certains dictionnaires incluant par exemple les dérivés comme mots-vedettes, tandis que d'autres ne les considèrent que comme des sous-entrées) et que les chiffres donnés pour les dictionnaires bilingues couvrent les deux sections (anglais-français et français-anglais).

– il s'agit de dictionnaires populaires, ou de grande renommée ; la notoriété du *(Nouveau) Petit Robert* et celle du *Petit Larousse* sont exemplaires. Pour les dictionnaires anglais, nous nous sommes basée sur les dictionnaires les plus vendus au moment du début de

[10] L'astérisque indique que la variété de langue mentionnée est celle qui est principalement décrite, même si d'autres variétés géographiques sont abordées.

[11] Les indications concernant la taille de la nomenclature sont celles fournies, telles quelles, par les diverses couvertures, préfaces et/ou introductions des dictionnaires en question.

l'étude (2002) à la librairie de l'Université d'Ottawa (Canada) et à la librairie Chapters d'Ottawa.

– il s'agit de dictionnaires dont le public visé correspond, globalement, à la catégorie suivante identifiée par Rey [2008 : 126] : « soit un locuteur natif dont la compétence est ressentie comme imparfaite, soit un apprenant de la langue décrite, à un niveau avancé » ; ce qui représente un public large sur le plan socioculturel, qui peut éventuellement « se définir par un niveau d'études (études secondaires complètes ou autodidactisme, par exemple) » —. ce qui est le cas notamment pour les versions "college" des dictionnaires américains.

• Critères propres aux dictionnaires unilingues :

– deux critères étaient en concurrence : le type de dictionnaire général (de langue ou encyclopédique) et la variété géographique de la langue décrite. Pour les dictionnaires français, nous avons choisi le premier critère et avons choisi d'une part un dictionnaire de langue (*NPR*) et d'autre part un dictionnaire encyclopédique (*PL*)[12]. Pour les dictionnaires anglais, c'est le deuxième critère qui a retenu notre attention, et nous avons donc décidé d'étudier deux dictionnaires présentant deux variétés d'anglais différentes : un présentant l'anglais nord-américain (*AH*), et un présentant l'anglais britannique (*NODE*), afin de pouvoir effectuer des comparaisons.

• Critère propre aux dictionnaires bilingues :

– nous avons choisi le dictionnaire *OXHA* parce qu'il repose explicitement sur un corpus, et le dictionnaire *HAR*, car celui-ci a la réputation de répertorier un grand nombre de mots scientifiques et techniques.

Passons maintenant à l'analyse croisée des données des corpus et dictionnaires que nous venons de décrire, tout d'abord concernant la macrostructure, puis concernant la microstructure.

3. Analyse de la macrostructure : corpus *vs*. dictionnaires

Notre analyse de la macrostructure des dictionnaires à la lumière des données du corpus se fait en deux temps. Dans un premier temps, nous effectuons une étude statistique globale afin de déterminer quel volume de notre liste de termes tirés du corpus est présent dans les dictionnaires. Dans un deuxième temps, nous nous intéressons à la représentativité de la nomenclature, c'est-à-dire à la présence ou l'absence de certains types de termes trouvés en corpus.

3.1. Remarques méthodologiques préalables

Concernant l'analyse de la macrostructure, il nous semble impossible de livrer des chiffres bruts sans expliquer au préalable dans quelles conditions le dénombrement s'est effectué. Comme nous l'avons expliqué plus haut, comptabiliser les termes présents dans un dictionnaire de langue générale est fort complexe. Nous nous sommes ainsi heurtée à de nombreuses difficultés que nous synthétisons et illustrons brièvement maintenant, et qui étaient principalement de trois ordres : la première concerne la forme du mot-vedette, la deuxième a trait au problème de la polysémie, enfin la dernière concerne l'emplacement du terme dans la microstructure du dictionnaire.

[12] Notons toutefois que la *Préface* du *PL* de 2002 décrit celui-ci comme « un dictionnaire de la langue française, un dictionnaire encyclopédique et un dictionnaire visuel ».

3.1.1. Forme du mot-vedette

Lorsque nous avons établi la liste des termes à analyser, nous avons constaté que certains termes trouvés en corpus présentent des formes concurrentes, encore appelées variantes, de divers types[13]. Il a fallu prendre en compte ces problèmes de variante lors du décompte des termes dans la nomenclature des divers dictionnaires. Nous n'avons bien évidemment comptabilisé les termes qu'une seule fois, même s'ils apparaissaient dans les dictionnaires sous diverses formes.

Certains termes présentent des variantes (ortho)graphiques : il s'agit principalement d'emprunts à d'autres langues (par exemple, *caldeira* / *caldera* en français, *aa* / *a'a* en anglais) ; d'autres variantes sont liés à des problèmes de nombre (*lapilli / lapillis* en français, où la désinence plurielle en *-i* de la forme italienne n'est plus perçue comme telle) ; d'autres encore peuvent être liées à la variété géographique (*dyke* (US) / *dike* (GB)) ; enfin, certaines variantes orthographiques sont liées à des problèmes de variation morphologique sur la racine du terme (*volcanologie vs. vulcanologie*).

Certains termes présentent des variations flexionnelles. Il s'agit pour nous de variations sur le nombre du terme : dans la liste des termes retenus, nous avons recensé les termes sous leur forme la plus couramment rencontrée dans le corpus (par exemple, *cendres* se trouve dans la plupart des cas au pluriel). Nous avons considéré qu'un terme était répertorié dans un dictionnaire même s'il se présentait sous une forme différente de celle retenue pour la liste initiale (par exemple, même si *cendre* est présenté au singulier dans un dictionnaire, comme le sens de volcanologie est mentionné, nous l'avons considéré comme répertorié).

Le troisième cas de variation rencontré était d'ordre morphosyntaxique. Il est représenté par l'alternance des parties du discours, notamment l'alternance nom / verbe[14]. Nous avons décidé de considérer un verbe comme non répertorié lorsque se présentait uniquement une forme nominale lui correspondant. C'est ainsi que, bien que *PL* répertorie le nom *émission* avec un sens de volcanologie, comme le montre l'entrée qui suit :

4. [Géologie] Sortie hors d'un volcan, lors d'une éruption, de produits magmatiques solides, liquides ou gazeux.

Figure 4. Extrait de l'entrée *émission* du *PL*

Nous n'avons pas considéré *émettre* comme répertorié, puisque l'entrée consacrée elle-même au verbe ne retenait pas le sens volcanologique. Nous avons appliqué la même politique, par exemple, pour le duo *réveil / se réveiller*, ainsi que pour *cool / cooling* ou encore *monitor / monitoring* dans divers dictionnaires.

Un autre cas problématique était celui de la présence ou de l'absence d'un élément d'une forme complexe. Ce problème de forme s'est posé de manière récurrente pour les adjectifs. En effet, même si un adjectif dispose d'une entrée propre, il n'est pas toujours défini ; parfois, il est défini uniquement à travers un composé. Nous n'avons considéré comme répertoriés que ceux d'entre eux qui présentaient une définition, comme c'est le cas pour *hawaïen* dans le *NPR*.

[13] La catégorisation des types de variation qui suit s'appuie sur celle établie par L'Homme [2004 : 74-75].
[14] Problème étudié tout particulièrement par Bourigault & Condamines [1999].

hawaïen, ïenne [...] **adj.** et **n.**
• 1839; de *Hawaï,* nom de la grande île d'un archipel du même nom
 Des îles Hawaï. — Spécialt *Guitare* hawaïenne. Chemise hawaïenne :* chemisette à col tailleur ornée de gros motifs floraux.
 Géol. *Volcan hawaïen,* dont l'édifice est surbaissé et la lave très fluide.

Figure 5. Extrait de l'entrée *hawaïen* du *NPR*

Enfin, dernier cas de figure de variation où le décompte est loin d'être simple : celui où coexistent deux formes, la forme simple et la forme complexe. Lors du dépouillement du corpus, nous avons été confrontée au problème des formes terminologiques complexes qui se réduisent au fil du discours. Le problème s'est posé tout particulièrement avec tous les noms simples qui entrent en cooccurrence avec l'adjectif *volcanique/volcanic*. Ainsi, nous avions éprouvé de la difficulté à décider si le terme à retenir pour notre liste initiale était *bombe* ou bien *bombe volcanique*. Le même problème s'est posé lors du dépouillement des dictionnaires : alors que nous avions retenu *bombe* comme terme initial à analyser, fallait-il considérer *bombe* comme répertorié lorsqu'il ne se présentait que sous la forme développée *bombe volcanique* (comme c'est le cas par exemple dans le *PL*) ? Nous avons considéré qu'il le fallait, et avons opéré de même, par exemple, pour *cône* et *cône volcanique*. En bref, nous avons considéré comme présents les syntagmes dont la tête (ou le déterminé) était répertoriée. Le problème se complexifie encore lorsqu'un terme simple peut être polysémique, et que la polysémie ne s'exprime qu'à travers la forme complexe. Par exemple, le terme *pahoehoe* (en anglais comme en français, d'ailleurs) peut désigner à la fois un type de lave, et un type de coulée de lave. Lorsqu'il s'agit de la première acception, les formes *pahoehoe* et *pahoehoe lava* sont en concurrence linguistique. Dans un des dictionnaires unilingues, seule paraissait la forme *pahoehoe lava* : nous avons considéré *pahoehoe* comme répertorié, même si un seul des sens était noté. Ceci nous amène à évoquer le problème de la polysémie de manière plus large.

3.1.2. Polysémie

Il a fallu effectuer un choix pour un problème autre que la variation, et qui s'est avéré plus que récurrent. Nombre de termes de la volcanologie partagent un signifiant commun avec des mots de la langue générale. Dans certains cas, la différence au niveau du signifié est claire : ainsi, le sens d'un mot de la langue générale comme *cheminée* est très nettement différent du sens volcanologique de *cheminée* (partie du volcan). Mais dans beaucoup d'autres, le sens d'un mot de la langue générale peut être étendu au domaine de la volcanologie sans qu'une véritable polysémie se dégage nettement. Après moult hésitations, nous avons décidé de ne pas considérer comme répertoriés les termes dont le sens était défini de manière trop générale, sans aucune allusion explicite à la volcanologie. Citons quelques exemples pour illustrer notre propos. Nous avons considéré comme absent de la nomenclature le terme *s'épancher* dans le *HAR*, le *NPR* et le *PL*. L'entrée du *NPR* est assez parlante à cet égard :

1□ Vx ou poét. Couler, se déverser. *« Je sentais comme une fontaine de miséricorde qui s'épanchait du haut du ciel dans mon cœur »* (Flaubert). — Mod. Méd. Former un épanchement. □ s'extravaser. *Sang qui s'épanche dans le cœur.*
□ Fig. Se répandre abondamment. « L'amour est inépuisable; □□□ plus il s'épanche, plus il surabonde » (Lamennais).

Figure 6. Extrait de l'entrée *s'épancher* du *NPR*

Notre choix peut d'autant plus se justifier par le fait que d'autres domaines sont mentionnés dans de telles entrées au détriment de celui de la volcanologie, comme c'est le cas dans l'exemple que nous venons de mentionner, où le sens médical de *s'épancher* est répertorié[15]. Notre choix peut également se justifier lorsque l'on compare un dictionnaire général à un autre. Par exemple, nous avons hésité devant l'entrée *bouchon* dans le *NPR*, dont une partie se présente comme suit :

4□ Par ext. Ce qui bouche accidentellement un conduit, un passage. *Bouchon de cérumen.*

Figure 7. Extrait de l'entrée *bouchon* du *NPR*

Or, la recherche de l'équivalent anglais de *bouchon* (*plug*) dans le *NODE*, donne comme sous-sens de *plug "*a mass of solidified lava filling the neck of a volcano". Cela nous a confirmé que nous ne pouvions considérer *bouchon* comme étant répertorié dans le *NPR*. Nous avons appliqué la même politique, par exemple, pour *dormir* dans le *NPR*, pour *repos* dans le *NPR*, et, surtout, pour *colère*[16] dans le *NPR* :

3□ Relig. et littér. La colère céleste, la colère divine. Jour de colère (cf. Dies iræ). — Poét. La colère des éléments, des flots. □ déchaînement.

Figure 8. Extrait de l'entrée *colère* du *NPR*

Le problème de la polysémie ne se pose pas qu'avec la langue générale. Il existe aussi des cas de polysémie avec des termes de domaines proches, soit superordonnés, soit connexes.

[15] Mentionnons que l'acception « couler » du verbe *s'épancher* est notée comme « littéraire » ou « obsolète » dans le *PL*.

[16] Le cas de *colère* mérite une explication. Une des grosses difficultés rencontrées lors du dépouillement des listes de candidats-termes a été de faire la différence entre métaphores de discours (à exclure) et métaphores lexicalisées (à conserver). Dans le contexte suivant trouvé en corpus *« Comme nous, un volcan naît, vit, s'endort, se réveille, est actif, se repose, fait des colères — les éruptions —, puis vieillit et finit par mourir »*, la présence d'un marqueur explicite de comparaison (*comme*) pourrait faire penser que l'ensemble des mots qui suivent sont des mots de la langue générale. Or certaines des métaphores utilisées ici sont clairement des métaphores lexicalisées qui sont des termes du domaine de la volcanologie : *se réveiller, (être) actif* ; qu'en est-il donc de *colère ?* Le contexte que nous venons de citer indique clairement que *colère* est synonyme d'*éruption*, et dans ce cas précis il s'agit d'une reformulation destinée à rendre plus accessible la compréhension des phénomènes volcaniques. Cependant, l'analyse des 43 occurrences de ce terme dans le corpus révèle que, dans 32 cas, *colère* est véritablement interchangeable avec *éruption* (*colère du Vésuve, colère du Pinatubo, colère de l'Etna*, *volcan en colère*…). Il semble donc que, même s'il est bien évident que *colère* n'est pas un terme de volcanologie à proprement parler, il faille retenir ce terme dans une optique de lexicographie générale. Mais la présentation de ce sens dans l'entrée d'un dictionnaire général est peut-être compliquée à mettre en place : puisque la présentation actuelle du *NPR* traitant implicitement le sens de volcanologie par le biais d'un sens principal (les sens secondaires n'étant pas forcément tous explicités) n'est pas satisfaisante, faut-il, dans l'entrée *colère*, prévoir une rubrique orientée vers une synonymie avec *éruption* ?

Parmi les termes qui figurent dans notre liste, certains relèvent plutôt du vocabulaire technique ou scientifique général. Lorsque le sens est strictement le même que celui utilisé en volcanologie, comme c'est le cas par exemple pour *dégazage* dans le *NPR*, nous avons considéré le terme comme répertorié. Le problème est plus épineux lorsque l'on évoque le domaine directement superordonné à la volcanologie, à savoir la géologie. En effet, même si plusieurs termes de la volcanologie partagent un signifiant commun avec des termes de la géologie, ce n'est pas pour autant que le signifié est le même, et la présence d'une marque de domaine *géologie* n'est pas suffisante pour conclure que le terme de volcanologie est répertorié[17]. L'exemple de *deflation* dans le NODE est assez révélateur à ce sujet :

3. *Geology* the removal of particles of rock, sand, etc. by the wind.

Figure 9. Extrait de l'entrée *deflation* du *NODE*

Dans certains cas, c'est bien uniquement le sens propre à la géologie qui est répertorié, mais sans qu'aucune marque ne laisse deviner qu'il s'agit de ce sens et non du sens propre à la volcanologie: c'est le cas de *pipe* dans le *NODE*, comme le montre l'extrait suivant où ne figure pas le sens propre à la volcanologie de « cheminée d'un volcan ».

pipe ***noun***
1. a tube of metal, plastic, or other material used to convey water, gas, oil, or other fluid substances.
cylindrical vein of ore or rock, especially one in which diamonds are found

Figure 10. Extrait de l'entrée *pipe* du *NODE*

Enfin, certains termes de volcanologie partagent un signifiant avec des domaines connexes : c'est le cas d'*aiguille*, qui possède un sens rocheux en géographie (mentionné dans le *PL* et le *NPR*). Or une aiguille, en volcanologie, est une aiguille de lave. Nous avons donc considéré *aiguille* comme absent dans ces deux dictionnaires.

Passons au dernier problème méthodologique auquel nous avons été confrontée pour le comptage des termes : celui de l'emplacement du terme.

3.1.3. Emplacement du terme

Comme son nom l'indique, la recherche en texte intégral implique que le terme est recherché dans n'importe quel endroit du dictionnaire, ce qui peut être un inconvénient. Notamment, la recherche nous indique parfois, dans les dictionnaires unilingues, qu'un terme est présent seulement dans la définition d'une entrée qui n'est pas le terme en question. Regardons l'exemple du terme *dynamisme* dans le *PL*. Alors que ce terme est utilisé de manière récurrente dans les définitions des adjectifs *péléen, hawaïen*, etc. (comme en témoigne l'entrée du *PL* qui suit), il n'est lui-même ni répertorié comme mot-vedette, ni défini dans son sens de volcanologie. Nous avons donc décidé de le considérer comme absent.

[17] Ceci montre en quoi le choix parfois fait par les lexicographes d'utiliser la marque du domaine superordonné peut être problématique. Il est facile pour un utilisateur de dictionnaire non initié aux domaines de la volcanologie et de la géologie de s'y perdre.

péléen adjectif, féminin **péléenne,** ou **peléen,** féminin **peléenne,** (de la montagne *Pelée*) [Géologie] Éruption ou dynamisme[18] péléens, éruption ou dynamisme caractérisés par l'émission de laves très visqueuses, sous forme de dômes ou d'aiguilles, et par des explosions très violentes provoquant la formation de nuées ardentes.

Figure 11. Extrait de l'entrée *péléen* du *PL*

3.2. Décompte global

Nous venons d'exposer divers obstacles méthodologiques qui se sont dressés sur notre chemin lors du comptage des termes de volcanologie répertoriés dans les dictionnaires généraux, et avons détaillé la manière dont nous avons tenté de les surmonter. Dans l'ensemble, nos critères d'inclusion sont plutôt stricts. Par conséquent, le nombre de termes est moins conséquent que ce qu'il aurait pu être si les critères avaient été plus larges. Il faut garder ce fait à l'esprit lorsqu'on observe les résultats du décompte.

Pour présenter les résultats de notre recherche, nous comparons tout d'abord les dictionnaires généraux entre eux, en comparant les données des dictionnaires unilingues entre eux, puis celle des dictionnaires bilingues entre eux, avant d'établir une comparaison entre dictionnaires unilingues et bilingues.

3.2.1. Comparaison des dictionnaires généraux unilingues entre eux

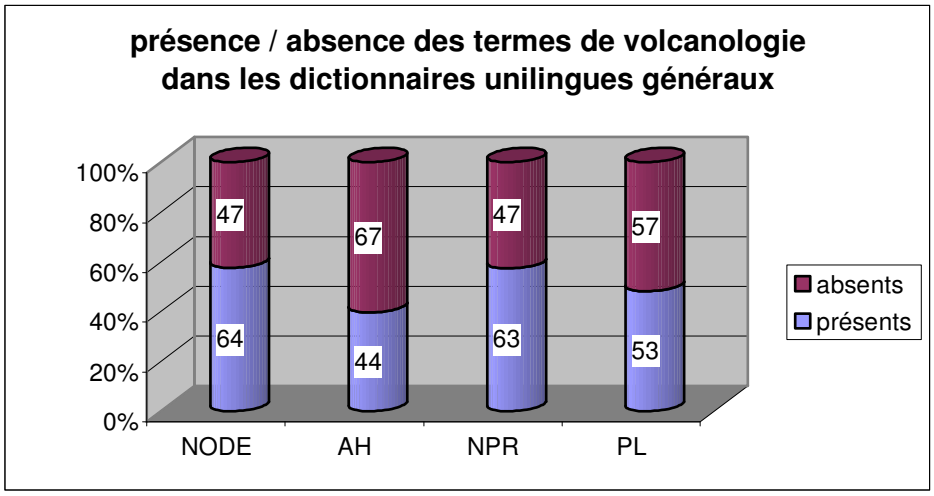

Figure 12. Présence / absence des termes de volcanologie dans les dictionnaires unilingues généraux analysés

Les données présentées dans le graphique ci-dessus invitent à faire trois types de comparaisons : tout d'abord une comparaison des dictionnaires unilingues français entre eux, puis une comparaison des dictionnaires unilingues anglais entre eux, enfin, une comparaison entre dictionnaires unilingues anglais et français.

• Dictionnaires français entre eux (*NPR* vs *PL*)

Quantitativement parlant, le *NPR* couvre mieux le domaine de la volcanologie que le *PL*, ce qui peut être étonnant étant donné qu'il est généralement reconnu qu'un dictionnaire encyclopédique (comme le *PL*) inclut plus de termes qu'un dictionnaire de langue, d'autant

[18] Souligné par nos soins.

que si l'on en croit les couvertures de ces deux dictionnaires (cf. tableau 3 où sont présentés les dictionnaires), l'étendue de la nomenclature est similaire.

• Dictionnaires anglais entre eux (*NODE* vs *AH*)

La différence entre les dictionnaires anglais est plus frappante que celle existant entre les dictionnaires français. Le dictionnaire britannique (*NODE*) paraît bien plus complet que le dictionnaire américain (*AH*), ce qui peut surprendre un peu, étant donné la compétence américaine en matière de volcanologie. Toutefois, souvenons-nous que l'*AH* est un dictionnaire de type 'college', donc prioritairement destiné aux étudiants et fait pour répondre à leurs besoins.

• Dictionnaires français vs dictionnaires anglais

Aucune différence frappante ne ressort de la comparaison entre dictionnaires français et dictionnaires anglais. Il semble possible de dire que, quantitativement parlant, ils en sont à peu près au même point. Globalement, le taux d'inclusion (c'est-à-dire la proportion de termes effectivement présents par rapport aux termes cherchés, présentée sous forme de pourcentages) avoisine les 50%, ce qui peut paraître satisfaisant. Toutefois, ce chiffre est à manier avec précaution : rappelons que nous avions choisi initialement le nombre de 110 termes à analyser parce que ce chiffre représentait une moyenne du nombre de termes de volcanologie inclus dans les dictionnaires généraux d'après une première recherche grossière ; or, après recherche approfondie des termes trouvés en corpus, aucun dictionnaire n'atteint ce chiffre : il y a donc nécessairement d'autres termes de volcanologie présents dans les dictionnaires généraux qui ont échappé à l'extraction de termes dans le corpus. Toutefois, si l'on classe les dictionnaires par ordre d'importance numérique du nombre de termes de volcanologie inclus, le classement est le suivant : *NPR* et *NODE* en tête, puis *PL*, puis *AH* nettement derrière.

3.2.2. Comparaison des dictionnaires généraux bilingues entre eux (OXHA vs. HAR)

	OXHA				*HAR*			
	F-E		E-F		F-E		E-F	
Termes français	36/110	**33 %**	30/110	**27 %**	30/110	**27 %**	24/110	**22 %**
Termes anglais	28/111	**25 %**	36/111	**32 %**	21/111	**19 %**	29/111	**26 %**

Tableau 4. Présence / absence des termes de volcanologie dans les dictionnaires bilingues généraux analysés

Aucune différence significative ne se dégage de la comparaison des deux dictionnaires bilingues analysés. Globalement, le taux d'inclusion va de 20 à 30%. Cela met un peu à mal l'image du *HAR*, qui est généralement reconnu pour son traitement privilégié des termes scientifiques et techniques. Au niveau quantitatif, il ne semble pas y avoir de différence particulière de traitement entre les deux dictionnaires, ou entre deux sections à l'intérieur d'un même dictionnaire. En revanche, ce que le tableau ne fait pas ressortir est le problème de cohérence des types de termes entre deux sections.

3.2.3. Dictionnaires unilingues vs. dictionnaires bilingues

La comparaison entre dictionnaires unilingues et dictionnaires bilingues fait ressortir, comme prévu, que le taux d'inclusion des termes des dictionnaires unilingues est supérieur à celui des dictionnaires bilingues : alors que les dictionnaires unilingues répertorient environ la moitié des termes recherchés, les dictionnaires bilingues n'en répertorient, grossièrement, qu'un tiers.

L'analyse quantitative de la nomenclature est une première approche intéressante, mais elle nécessite d'être complétée par une analyse de type qualitatif, que nous présentons maintenant.

3.3. Analyse selon le type de terme

En nous basant sur les études métalexicographiques existantes qui énumèrent un certain nombre de critères potentiels d'inclusion des termes (par exemple, Boulanger & L'Homme [1991], Thoiron [1998], ou Candel & Humbley [1997]) et sur les données propres au domaine de la volcanologie, nous avons utilisé les critères suivants pour l'analyse des divers types de termes dans la nomenclature : appartenance à une terminologie organisée, métaphorisation, morphologie, catégorie grammaticale, aspect culturel, variation.

3.3.1. Appartenance à une série conceptuelle

La question que le lexicographe chargé des termes dans un dictionnaire général est en droit de se poser est la suivante : la terminologie d'un domaine doit-elle être présentée dans un dictionnaire général en respectant scrupuleusement l'arbre du domaine? Selon Alain Rey [1985 : 5], le dictionnaire encyclopédique se doit de le faire, tandis que cela a moins d'importance pour le dictionnaire de langue :

> Alors que le dictionnaire encyclopédique, comme le recueil strictement terminologique, se doit de distribuer les termes selon un arbre de domaines bien construit, le dictionnaire de langue doit marquer la nature linguistique du terme, lequel appartient à un registre d'usage marqué (comme technique, scientifique, didactique, et éventuellement par une marque plus précise – nom d'une technique ou d'une science), mais dont la finalité n'est pas classificatoire quant au contenu notionnel. Les classifications du dictionnaire de langue sont – et doivent être – d'abord et essentiellement linguistiques, voire sociolinguistiques.

Une question qui découle de la précédente est donc : faut-il inclure tous les termes qui se situent à un même niveau hiérarchique dans la taxonomie, respectant en cela la démarche proprement terminologique, même si cela ne reflète pas l'importance desdits termes dans la langue générale? C'est le problème que soulève Béjoint [1988 : 361] :

> One would think that the coverage of each domain could be regulated by the taxonomic organization of scientific and technical words. The lexicographer would draw up a complete taxonomy of all the words of each domain, and record words 'layer by layer', taking care that all the words of the same hierarchical level are added, or substracted, together. But this is rendered difficult by the co-existence of competing taxonomies in certain domains. And even when the taxonomies are undisputed (as in the classification of living things), the place of

> scientific and technical words in their taxonomies does not correspond to their order of importance in the general lexicon.

Certains métalexicographes, comme Thoiron [1998 : 628 *sq*.], font ainsi l'évaluation des terminologies présentes dans les dictionnaires à la lumière de la structuration du domaine. Celui-ci explique notamment qu'il est « commode de repérer l'absence des co-hyponymes ou celle des hyperonymes », disant que « c'est probablement sur ces critères qu'on peut le mieux évaluer la cohérence conceptuelle d'une macrostructure [...] en se souvenant toutefois que la fréquence d'un terme n'est pas corrélée avec sa position dans une taxinomie » (*ibid.*).

Avec cet exemple en tête, il nous a donc semblé important d'utiliser le critère de l'appartenance à une série conceptuelle bien délimitée pour évaluer la macrostructure des dictionnaires, en étudiant des hyperonymes, des hyponymes et co-hyponymes. Une analyse très détaillée de ce critère, reposant sur l'analyse de deux séries conceptuelles (celle des projections volcaniques et celle du degré d'activité du volcan) a déjà été présentée dans Josselin-Leray & Roberts [2007]. Nous n'en livrerons ici que les conclusions.

Deux séries conceptuelles ont été choisies ; elles l'ont été parce que l'ensemble des termes qui les composent présentent un nombre élevé d'occurrences dans le corpus (par exemple, *active volcano* apparaît 242 fois dans la partie anglaise du corpus comparable), et parce qu'elles sont très différentes l'une de l'autre. Dans la première série, celle des projections volcaniques, les référents sont concrets : il s'agit de matériaux émis par le volcan qui peuvent être touchés, mesurés, et dont la composition peut être analysée, ce qui semble assez simple à décrire. Dans la deuxième série, celle du degré d'activité d'un volcan, les référents sont abstraits et construits par l'homme de façon assez arbitraire, ce qui cause un certain nombre de problèmes d'ordre sémantique. L'analyse macrostructurelle de deux types de séries conceptuelles nous a livré les informations suivantes : pour la première série, celle des projections volcaniques, il semble que le critère du degré d'opacité du terme primerait sur le critère de fréquence pour l'inclusion d'un terme dans un dictionnaire général (unilingue, du moins) ; pour la seconde série, celle du degré d'activité du volcan, les collocations anglaises semblent mieux traitées. Cela s'explique peut-être par le fait qu'il n'y a pas de phénomène de synonymie (ou de variation) en anglais pour cette série. La plupart des collocations françaises (et tout particulièrement *volcan actif*) ne reçoivent pas autant d'attention qu'elles le mériteraient dans les dictionnaires si l'on se fie aux données du corpus. Il est difficile de tirer des conclusions sur les critères d'inclusion des termes utilisés dans les dictionnaires français, puisque ni la fréquence du terme dans les textes de volcanologie, ni l'appartenance du terme à un système notionnel bien délimité ne semblent être les critères prédominants.

3.3.2. Termes métaphoriques

Un autre critère que les lexicographes disent utiliser est le suivant : sera retenu un terme qui donne lieu à une métaphore dans la langue générale. C'est ce qu'expliquent les auteurs de la septième édition du *Dictionnaire de l'Académie française* (Quemada [1997 : 403]), et ce que recommande également Cottez [1994 : 20] :

> 3^{e} critère d'admission : [Le terme] commence à se prêter à un emploi métaphorique ou figuré. C'est le cas de freiner, qu'on ne peut éliminer en ce début du 20^{e} siècle parce qu'il apparaît dans toutes sortes de comparaisons très naturelles. Ce qui signifie que le cordon ombilical qui le relie à son emploi originel est rompu : il a acquis son autonomie et va mener sa vie propre.

Le phénomène de métaphorisation n'a pas lieu que dans un seul sens : il arrive fréquemment que des termes soient des métaphores de mots de la langue générale. Ceux-ci sont impérativement à répertorier, d'après Cottez [ibid. : 22].

Qu'en est-il dans notre domaine ? Afin d'effectuer cette évaluation, nous avons regroupé ici tout d'abord les données relatives aux métaphores liées à l'idée de sommeil, présentes surtout en français. Nous nous contentons ici de livrer les résultats concernant les verbes *dormir* et *se réveiller*, ainsi que ceux concernant les substantifs *réveil, sommeil, repos*. Nous incluons également le terme *colère*, dont nous avons déjà parlé précédemment (cf. note 15).

	Nb occ.[19]	*NPR*	*PL*	*OXHA (F-E)*	*HAR (F-E)*
réveil	70	+	–	+	+
sommeil	44	–	–	–	–
se réveiller	78	–	–	+	–
dormir	37	–	–	–	–
repos	18	–	–	–	–
colère	43	–	–	+	–

Tableau 5. Présence / absence des termes métaphoriques français dans les dictionnaires

Les dictionnaires unilingues français répertorient très peu ces usages métaphoriques (le *PL* n'en répertorie même aucun). Or, rappelons que *colère* était un terme fréquent dans la partie française du corpus comparable puisqu'il présentait 43 occurrences. Ce sont finalement les dictionnaires bilingues qui semblent les plus prêts à accueillir ce genre de terme dans leur nomenclature, sans doute parce qu'il est toujours difficile de savoir si une métaphore dans une langue se traduit par la même métaphore dans une autre langue.

3.3.3. Termes morphologiquement dérivés

Certains auteurs, comme Cottez [1994], revendiquent la (possibilité de) dérivation d'un terme comme critère d'inclusion. Qu'en est-il dans les faits ?

En s'intéressant à la dérivation entre nom et adjectif, Thoiron [1998 : 632] cherche à établir s'il y a une « cohérence du point de vue linguistique » dans la macrostructure du dictionnaire qu'il analyse, l'*OXHA*. Il note l'importance de ce critère d'un point de vue terminologique :

> C'est un aspect important en terminologie puisque la sous-catégorisation implique souvent l'ajout d'un composant adjectival.

Pour l'analyse des termes dérivés dans notre domaine, nous nous contentons d'analyser l'exemple de l'adjectif français *volcanologique* et de son équivalent anglais *volcanological*. Ces termes peuvent en effet être considérés comme importants pour la terminologie du domaine, et, de plus, le traitement qui leur est réservé est emblématique du traitement des dérivés dans les dictionnaires (qu'ils soient généraux ou spécialisés par ailleurs).

[19] Dans ce tableau et tous ceux qui suivent, sauf mention explicite contraire, le nombre d'occurrences indiqué (parfois appelé fréquence en référence à la fréquence brute) est celui trouvé dans le corpus comparable, non dans le corpus traduit.

Nb occ.	*NPR*	*PL*	*OXHA (F-E)*	*HAR (F-E)*
50	+	+	–	–

Tableau 6. Présence / absence du terme *volcanologique* dans les dictionnaires

Nb occ.	*NODE*	*AH*	*OXHA (E-F)*	*HAR (E-F)*
7	+	–	–	–

Tableau 7. Présence / absence du terme *volcanological* dans les dictionnaires

Le phénomène est frappant : très peu de dictionnaires répertorient ces dérivés ; notamment, aucun dictionnaire général bilingue ne le fait. Seuls trois dictionnaires unilingues les incluent : *volcanologique* est présent dans le *PL* et le *NPR*, tandis que *volcanological* n'est répertorié que dans *NODE*. Alors que l'absence du terme anglais pourrait se justifier par la faible fréquence en corpus, il n'en va pas de même pour le terme français. Toutefois, un parcours rapide des contextes du corpus permet de voir que *volcanologique* est principalement utilisé dans le composé *observatoire volcanologique*. La transparence du terme (anglais ou français) pourrait également expliquer son absence des nomenclatures.

3.3.4. Catégorie grammaticale

Les résultats de l'analyse faite par Mazière [1981 : 91] montrent que les dictionnaires généraux reflètent en partie la terminologie telle qu'elle est vue par les terminologues. Or, comme l'a souligné L'Homme [2002b] pour les dictionnaires spécialisés, les catégories autres que le nom, telles que l'adjectif et le verbe, suscitent de l'intérêt chez les terminologues depuis peu. Nous avons donc examiné comment les termes répertoriés dans les dictionnaires généraux étudiés se répartissent selon la catégorie grammaticale : adjectif, verbe, nom simple ou syntagme nominal.

En ce qui concerne les adjectifs (au nombre de 16 dans notre liste), ceux-ci sont plus durement traités dans les dictionnaires bilingues que dans les dictionnaires unilingues généraux. Les dictionnaires unilingues qui répertorient proportionnellement plus d'adjectifs sont le *NODE* (81%) et le *NPR* (69%). Quant aux dictionnaires bilingues, il ne semble pas y avoir de différence quantitative flagrante, d'une part, entre dictionnaires, et, d'autre part, entre sections d'un même dictionnaire : le taux d'inclusion avoisine les 30% pour les termes français, et est légèrement plus élevé pour les termes anglais.

Pour les verbes (au nombre de 6), le bilan est bien maigre : le verbe spécialisé (comme *dormir*), ou le verbe entrant dans une collocation verbale (comme *monitor a volcano*) est bien le parent pauvre des dictionnaires, et ce quelle qu'en soit la langue ou quel qu'en soit le type : 0% dans les unilingues français, 33% dans les unilingues anglais ; un taux d'inclusion qui varie entre 17 et 33% pour les bilingues).

Passons à la catégorie nominale. Les noms simples (au nombre de 52) sont plutôt bien représentés dans les dictionnaires généraux unilingues. De nouveau, ce sont les dictionnaires unilingues *NPR* et *NODE* qui en incluent proportionnellement le plus grand nombre. La tendance générale est que les dictionnaires bilingues (de 27 à 46%) répertorient moins de noms simples que les dictionnaires unilingues (de 54 à 68%). Le dictionnaire *OXHA* semble inclure légèrement plus de substantifs que le *HAR*, mais on ne remarque aucun écart majeur.

Pour ce qui est des syntagmes nominaux (au nombre de 36), les résultats viennent contrecarrer l'idée fort répandue selon laquelle les dictionnaires anglais répertorient plus de termes complexes ou syntagmes nominaux que les dictionnaires français (dans notre cas, *NPR* : 50%, *PL* : 39%, *NODE* : 43%, *AH* : 19%). L'*AH* fait particulièrement piètre figure à ce sujet. Les chiffres vont aussi à l'encontre de l'idée selon laquelle les dictionnaires encyclopédiques incluraient plus de termes complexes que les dictionnaires de langue. En effet, le *NPR* et le *NODE* (dictionnaires de langue) recensent proportionnellement plus de syntagmes que le *PL* et l'*AH* (dictionnaires encyclopédiques). Enfin, un dernier résultat est frappant : la place accordée aux syntagmes nominaux dans les dictionnaires bilingues est plus que réduite, puisqu'elle ne s'élève jamais au-dessus de 19%.

3.3.5. Aspects culturels

En constituant nos corpus de vulgarisation, nous avions pu constater la chose suivante : les articles sur la volcanologie apparaissent avec une fréquence totalement irrégulière dans les médias, et les centres d'intérêt sont très liés à la culture qui les produit. Autrement dit, les journaux français vont essentiellement parler des volcans de France (ce qui peut inclure donc, outre la chaîne des Puys en Auvergne, les volcans des Dom-Tom tels que le Piton de la Fournaise à la Réunion) et d'Italie, tandis que les journaux britanniques s'intéresseront aux volcans de l'île de Montserrat et que les journaux américains se pencheront surtout sur les volcans hawaiiens. Il nous a donc semblé intéressant de voir comment cet aspect « géographique » ou « culturel » de la terminologie présent dans le corpus était reflété dans les dictionnaires. Nous avons donc examiné des termes correspondant à des réalités françaises et des réalités américaines.

Ainsi, nous avons choisi de voir comment était représenté le terme *puy*, terme franco-français s'il en est, qui réfère en particulier aux volcans d'Auvergne. Il y a 174 occurrences de ce terme dans le corpus comparable.

NPR	*PL*	*OXHA (F-E)*	*OXHA (E-F)*	*HAR (F-E)*	*HAR (E-F)*
+	+	–	–	–	–

Tableau 8. présence / absence du terme français *puy* dans les dictionnaires généraux

Les dictionnaires unilingues ne peuvent faire l'impasse sur ce terme. L'un cite comme exemple le Puy de Dôme (*NPR*), et l'autre cite l'Auvergne dans sa définition (*PL*). En revanche, les dictionnaires bilingues font l'impasse sur ce terme, ce qui s'avère d'autant plus gênant qu'il mériterait un équivalent explicatif.

Pour ce qui est des réalités américaines, nous avons choisi de regarder si les termes *aa* (96 occurrences) et *pahoehoe* (87 occurrences), termes d'origine hawaïenne renvoyant à des réalités hawaïennes, donc, américaines, étaient répertoriés.

	NODE	*AH*	*OXHA (F-E)*	*OXHA (E-F)*	*HAR (F-E)*	*HAR (E-F)*
aa	+	–	–	–	–	–
pahoehoe	+	–	–	–	–	–

Tableau 9. présence / absence des termes anglais *aa* et *pahoehoe* dans les dictionnaires généraux

Il est fort surprenant que ce soit le *NODE*, dictionnaire britannique, et non l'*AH*, dictionnaire américain, qui répertorie ces deux termes. L'absence des termes dans les dictionnaires bilingues s'explique sans doute par le caractère spécialisé du terme (emprunt).

Passons au dernier type de terme dont la représentativité mérite d'être analysée : les variantes.

3.3.6. Variantes

Comme nous l'avons indiqué plus haut, l'analyse des données du corpus nous a permis de constater le très grand nombre de variantes possibles pour un même terme. L'analyse des divers types de variantes (graphique, flexionnelle, morphosyntaxique) dans les dictionnaires généraux étudiés fait ressortir diverses tendances.

Les variantes de type graphique sont assez bien représentées dans la nomenclature des divers dictionnaires ; c'est notamment le cas des variantes géographiques (comme *dyke / dike*). Toutefois, certaines lacunes sont à noter : ainsi, le français hésite sur la forme du mot *caldeira* / *caldera*. Le corpus a montré que les deux formes existent (117 occurrences de *caldeira* contre 47 de *caldera*), mais le *PL*, par exemple, ne signale que la forme la plus fréquente.

On note une correspondance entre données du corpus et données des dictionnaires sur les formes *volcanologue* et *vulcanologue*. Le corpus indique clairement que la forme *vulcanologue* est bien moins fréquente (35 occurrences par rapport à 290), ce que reflètent les dictionnaires en indiquant que cette dernière forme est désuète.

Certains termes présentent des variations flexionnelles qui concernent le nombre du terme. Pour cet aspect, on note une grande différence entre les données du corpus et les données des dictionnaires généraux. Ceci est dû au fait que les dictionnaires généraux présentent le mot-vedette sous la forme d'un lemme. Mais cela peut parfois mener à de véritables non-sens. Regardons la présentation du terme *lapilli* dans l'*AH* :

> **la·pil·lus** *n.*, *pl.* **la·pil·li**. A small, solidified fragment of lava. [Latin, diminutive of *lapis*, stone.]

Figure 13. Entrée *lapilli* du *AH*

Ce dictionnaire présente une forme (*lapillus*) qui n'est aucunement attestée dans le corpus comparable : le corpus ne présente que la forme *lapilli*. Cela s'explique sans doute par le fait que l'étymologie que l'*AH* donne de ce terme est erronée, ou du moins incomplète, si l'on en croit l'étymologie donnée par d'autres dictionnaires (comme le *PL*, par exemple) : ceux-ci indiquent que le terme *lapilli* est un terme italien (il s'agit donc sans doute d'un terme qui provient originellement du latin, mais qui est passé par l'italien). Le constat fait par Tetet [1994] concernant l'aspect lacunaire des informations étymologiques sur les termes se confirme donc.

Enfin, les divers cas de variation morphosyntaxique sont largement sous-représentés dans les dictionnaires généraux, cela sans doute pour un problème de place. Le cas le plus frappant est celui de toute la série des termes se présentant ou non en cooccurrence avec l'adjectif *volcanique* (ex. *bombe/bombe volcanique*). Alors qu'un dictionnaire unilingue spécialisé comme le *Dictionnaire de la tectonique des plaques et de la géodynamique* (de 2000) indique clairement la variation par l'usage de parenthèses autour de l'adjectif *volcanique* ou par l'utilisation de la conjonction *ou* (ex : *bombe* ou *bombe volcanique*), il n'en est fait mention que dans le *NODE* (on trouve ainsi, dans l'entrée *bomb*, « also volcanic bomb »). Du moins, il n'en est nulle part fait mention *systématiquement*, et la présentation peut varier à l'intérieur d'un même dictionnaire, ce qui nous amène maintenant à évoquer les problèmes de microstructure.

4. Analyse de la microstructure : corpus *vs*. dictionnaires

Sur quels aspects de la microstructure le recours au corpus peut-il avoir une influence décisive ? Nous nous concentrons ici sur les principaux, à savoir : les indications grammaticales (valable pour les deux types de dictionnaire), les définitions (valable pour les dictionnaires unilingues) et les équivalents (valable pour les dictionnaires bilingues).

4.1. Indications grammaticales

4.1.1. Structure des collocations : épithète et attribut

Reprenons l'exemple de la série conceptuelle désignant le degré d'activité d'un volcan (*volcan actif/active volcano*, etc.). Dans ces collocations de type adjectif + nom, l'adjectif peut être soit attribut, soit épithète. La comparaison des données du corpus comparable avec celles des dictionnaires généraux bilingues nous a permis de constater un défaut de présentation majeur. Pour *active volcano*, le *HAR* présente le schéma où *active* est attribut (**to be/become active (of volcano)** être / entrer en activité)) ; cela pourrait inciter un utilisateur du dictionnaire non-anglophone à croire que *active* fait partie de la série des adjectifs en *a-* ne pouvant être utilisés qu'en position d'attribut (*asleep, ajar,* etc.). Or, les données du corpus comparable nous ont montré que le schéma syntaxique le plus fréquent est celui où l'adjectif est épithète. Le problème est inversé pour l'adjectif *inactive* dans les deux dictionnaires bilingues : ceux-ci présentent l'adjectif en position d'épithète (*inactive volcano*), alors que les exemples tirés du corpus montrent que la position d'attribut est plus courante, comme en témoigne l'extrait suivant tiré du corpus comparable : *« a volcano which is presently* ***inactive*** *but which may erupt again »*.

4.1.2. Valence des verbes

Nous avons décidé d'analyser en détail le verbe *erupt*, étant donné, d'une part, que ce verbe était le plus fréquent de la partie anglaise du corpus comparable (pour mémoire, 554 occurrences) et, d'autre part, que le parcours de certains contextes de ce corpus nous avait alertée sur une éventuelle variation de construction syntaxique (transitive ou intransitive) ayant de potentielles répercussions sur la traduction, comme nous l'avons expliqué plus haut (cf. section 2.1.2). Nous avons donc étudié la répartition des deux constructions dans la partie anglaise du corpus comparable, puis avons décidé de poursuivre nos investigations dans la partie anglaise du corpus traduit. Les résultats chiffrés se trouvent dans le tableau qui suit :

	Construction intransitive	Construction transitive
Corpus comparable	512 / 544 (92 %)	42/ 544 (8 %)
Corpus traduit	43 / 44 (97,7 %)	1 / 44 (2,3 %)[20]

Tableau 10. Nombre d'occurrences en corpus des formes transitives et intransitives du verbe *erupt*

L'étude complète de ce verbe figure dans Frérot & Josselin-Leray [2008]. Nous présentons ici les principales conclusions auxquelles nous ont menée cette analyse.

[20] La seule occurrence de la construction transitive se trouve dans un texte-cible du corpus traduit (il s'agit donc d'une construction utilisée par le traducteur).

• **Analyse de la construction intransitive**

Dans le corpus comparable, parmi les 512 occurrences du verbe intransitif, 5 présentent le cas où le verbe *erupt* est utilisé dans un sens figuré et ont donc été écartées. Nous avons également trouvé 15 occurrences de la forme *erupting volcano*, qui ne sont pas comptabilisées dans le chiffre 512, mais qui devraient figurer dans les dictionnaires bilingues du fait que la traduction en français présente un schéma syntaxique différent du schéma syntaxique anglais : la forme française est du type nom + préposition + nom (*volcan en éruption*).

Le dépouillement du corpus traduit, quant à lui, livre 43 occurrences du schéma intransitif du verbe *erupt* (sous ses diverses formes : *erupt, erupts, erupted, erupting*), ce qui correspond à la quasi-totalité des occurrences.

L'analyse de tous les contextes où apparaissait le verbe *erupt* dans sa construction intransitive nous a permis d'établir que les sujets les plus typiquement utilisés avec le verbe *erupt* (dans 67% des cas pour le corpus comparable et 43,2% des cas pour le corpus traduit) sont *volcano(es)* ou un nom de volcan. Ceci, ainsi que le fait que la plupart des contextes proviennent de quotidiens généralistes, pourrait expliquer que l'on ne trouve que *volcano* comme sujet du verbe *erupt* dans les dictionnaires généraux analysés (cf. section 2.1.2). Toutefois, les autres sujets (*magma, lava, molten rock*...) ne devraient pas être mis de côté, car ils représentent une part non-négligeable des contextes répertoriés. De plus, utilisé avec ces autres sujets, *erupt* ne peut pas être traduit par *entrer en éruption*, qui est l'équivalent habituellement présenté par les dictionnaires bilingues. En effet, *magma, lava*, et les autres exemples relevés représentent, de manière générale, tout ce qui peut être « éjecté du volcan ». Cet usage est bien répertorié dans le *NODE* :

be ejected from an active volcano: *hot lava* ***erupted from the crust.***

Figure 14. Extrait de l'entrée *erupt* du *NODE*

• **Analyse du verbe transitif**

Le verbe *erupt* est utilisé de manière transitive dans 43 occurrences sur 554, soit dans environ 8% du nombre total d'occurrences du verbe. Ces occurrences proviennent de diverses sources, et correspondent aux trois schémas syntaxiques suivants[21] :

(i) X erupts Y

(5) Volcanoes on other planets, and their moons, do not necessarily erupt the same materials as Earth's volcanoes

Ce schéma est le schéma typique de l'usage transitif du verbe (corpus comparable : 11 occurrences, corpus traduit : 1 occurrence).

(ii) Y is erupted (from X / by X)

(6) Volcanic material can also be erupted **from** the side of a volcano to produce a flank eruption

Le deuxième schéma correspond à la forme passive du verbe (15 occurrences dans le corpus comparable).

(iii) Y erupted (from X / by X)

(7) Total volume of tephra erupted during the formation of Crater Lake was 18 cubic miles

Le troisième schéma syntaxique est le plus elliptique, utilisant seulement le participe passé[22]. Nous avons trouvé 18 occurrences de ce schéma dans le corpus comparable.

S'il est vrai que la construction transitive ne représente qu'environ 8% des occurrences du verbe dans notre corpus, il nous semble que les dictionnaires généraux devraient

[21] Les deuxième et troisième schémas syntaxiques sont en fait issus du premier.

[22] On peut aussi considérer ces constructions comme des propositions relatives réduites.

néanmoins la répertorier, et sous toutes ses formes : en ce qui concerne les dictionnaires unilingues, le sens est vraiment différent du sens le plus courant de *erupt* intransitif (*entrer en éruption*)[23] ; en ce qui concerne les dictionnaires bilingues, la traduction diffère de l'habituel *entrer en éruption*.

4.2. Définitions

4.2.1. Remarque préliminaire

Il nous a semblé intéressant de déterminer, parmi les termes présents dans les dictionnaires analysés, lesquels faisaient l'objet d'une définition.

NPR		*PL*		*NODE*		*AH*	
30/63	**48%**	43/53	**81%**	50/64	**78%**	38/44	**86%**

Tableau 11. Nombre de termes définis par rapport au nombre de termes répertoriés dans les unilingues

La comparaison de la proportion de termes définis vient relativiser les « scores » quantitatifs présentés dans l'analyse de la macrostructure. En effet, le *NPR* a beau inclure de très nombreux termes de la volcanologie, à peine la moitié d'entre eux sont définis (de nombreux termes ne sont ainsi présents que dans les exemples). A l'opposé, l'*AH*, qui avait un taux d'inclusion relativement faible, est celui qui définit le plus grand nombre de termes.

4.2.2. Précision et type

Le problème des définitions des termes dans les dictionnaires généraux a été largement traité dans la littérature existante**.** Les problèmes principaux tournent autour de trois axes majeurs : la précision linguistique et scientifique de la définition, la source des définitions, le type de définition. Nous nous intéressons ici au premier et au dernier.

Béjoint [1988 : 362 *sq*.] fait le constat suivant à la suite de l'analyse des définitions dans plusieurs dictionnaires unilingues anglais : un bon nombre de définitions présentent un caractère très vague, et incluent des adverbes très généraux comme « usually, generally, especially…etc. ». D'après lui, ce flou linguistique est relié à un flou notionnel, notamment dans le cas des mots qui entrent en vulgarisation (« words fairly common in non-specialized use »), et auxquels il est difficile d'accorder un sens précis. Il explique également le caractère flou de certaines définitions par le fait que les lexicographes se sentent obligés d'utiliser un vocabulaire simple, voire simpliste, pour les définitions. Mais le recours à un vocabulaire spécialisé serait tout aussi problématique, selon lui [*ibid.* : 357-358].

Les auteurs ayant étudié les définitions des termes évaluent également leur exactitude scientifique. Wesemael & Wesemaël [2003], par exemple, se posent la question suivante : « Une image juste de la science est-elle véhiculée par le dictionnaire général ? ». Béjoint [1988 : 364] explique que le degré de précision scientifique peut varier beaucoup d'un dictionnaire à l'autre. Il indique notamment que les dictionnaires américains ont souvent une approche plus scientifique de la définition que les dictionnaires britanniques. Il semble d'autre part que les dictionnaires français les plus récents versent beaucoup dans la « scientifisation » des définitions, comme le regrette Raimbault [2003 : 10].

[23] En revanche, il est assez proche du sens que l'on trouve dans les énoncés où le sujet de *erupt* intransitif est autre que *volcano* ou un nom de volcan.

Tout dépend en réalité du but que s'assigne le dictionnaire : véhiculer une image scientifique la plus juste qu'il soit, ou se mettre à la portée de l'utilisateur. L'idéal serait un juste milieu entre les deux. Citons un exemple bien connu pour éclaircir notre propos. Béjoint [1988 : 365] compare les deux phrases suivantes :

(8) You can't see the stars because of the sun
(9) This star is the sun

Pour un scientifique, le soleil est une étoile, ce qui rend la phrase (8) inacceptable pour lui, alors qu'elle est acceptable et compréhensible pour un utilisateur profane. La phrase (9), en revanche, est inacceptable pour un profane. Béjoint [1988] cite également le problème de la définition des couleurs, pour lesquelles les lexicographes oscillent entre une définition utilisant les longueurs d'onde et une définition comparant la couleur à un référent universel (ex. : le bleu est la couleur du ciel par une belle journée). Ce même auteur suggère de régler le problème du degré scientifique de la définition en distinguant le sens « commun » du sens « scientifique » et en écrivant deux définitions différentes, solution adoptée par le *NODE*.

Dans Josselin-Leray & Roberts [2007], nous avons longuement étudié la définition des termes des deux séries conceptuelles mentionnées en 3.1.3.1. Nous résumons ici en quelques lignes les principales conclusions. Les dictionnaires étudiés offrent les deux extrêmes pour ces termes : ainsi, pour le terme *volcan éteint* trouvé sous *éteint* dans le *NPR,* la définition présente une simplicité excessive, qui frôle l'inexactitude (« qui ne brûle, n'éclaire plus ») ; alors que certaines définitions font à l'inverse preuve d'une complexité qui peut laisser perplexe (le *AH* dit d'un *dormant volcano* qu'il est "temporarily quiescent"), ou d'une degré de précision qui peut paraître superflu (le *PL* définit le terme *lapilli* comme des « projections volcaniques de petites dimensions (entre 2 et 64 mm de diamètre) ». Les données du corpus nous ont permis de voir qu'entre ces deux pôles, celui d'une extrême simplicité et celui d'une extrême technicité, un juste milieu était envisageable. Par exemple, pour ce dernier terme, il nous a semblé qu'à défaut de donner des précisions millimétriques qui ne sont pas nécessairement indispensables dans un dictionnaire général, un compromis pourrait être trouvé entre classification dite granulométrique et approximations grossières ou inexistantes. C'est l'avantage que représente un corpus de vulgarisation comme celui que nous avons constitué. En effet, le contexte suivant, tiré de la partie française du corpus comparable, pourrait servir de base à l'élaboration d'une définition lexicographique plus fine :

(10) Les volcanologues classent les projections volcaniques selon leurs dimensions, des plus petites aux plus grosses :
- les « poussières », aussi fines que de la farine ;
- les « cendres » (n'ayant rien à voir avec celles d'une cigarette ou d'un feu), ressemblant à du sable ;
- les « lapilli » (« petites pierres », en italien), pouvant atteindre la taille d'une balle de tennis ;
- les « blocs », sans limite de taille.

Une telle description fait appel aux connaissances de base du grand public, à des référents familiers, et serait en même temps d'une plus grande précision que celle actuellement présente dans les ouvrages lexicographiques.

Le deuxième problème relatif à la définition des termes est le suivant : quel type de définition est, ou doit être, utilisé pour les termes dans les dictionnaires généraux ? Le premier constat que fait Béjoint [1988 : 364] est le suivant :

> Ordinary, context-based definitions of the type used for common words are clearly inadequate.

Toutefois, il explique qu'une vision traditionnelle de la définition, du type *genus* + *differentiae* paraît envisageable. Dans la vision idéale, le terme est remis à sa place dans la structure du domaine par le biais de la définition : le *genus* serait l'hyperonyme du terme, et les *differentiae* ses co-hyponymes. Mais un des problèmes est que tous les éléments de la taxonomie ne sont pas toujours répertoriés dans les dictionnaires (comme nous l'avons évoqué dans l'analyse de la macrostructure), notamment à cause de problèmes de fréquence (un hyperonyme est souvent moins fréquent). Notre confrontation des données du corpus avec celles des dictionnaires concernant les deux séries conceptuelles nous permet de faire les deux remarques suivantes : (1) même si le terme générique est absent ou non défini dans le dictionnaire, il est possible de le représenter systématiquement dans le dictionnaire sous la forme d'une paraphrase qui peut être trouvée en corpus (ex. : « fragment de lave projeté » pour *projections* ou *tephra*) ; (2) une analyse fine des données du corpus peut permettre de mieux identifier le(s) trait(s) sémantique(s) qui distinguent les co-hyponymes les uns des autres et donc d'améliorer la finesse de la définition (ex. : la différence entre *volcan éteint* / *volcan actif* / *volcan endormi* est constituée par la présence ou l'absence de l'un ou plusieurs des traits sémantiques suivants : « actuellement en éruption », « a connu des éruptions au cours de la période historique », « susceptible de se réveiller »).

Nous avons consacré cette sous-partie à un problème propre aux dictionnaires unilingues. Intéressons-nous à présent aux dictionnaires bilingues.

4.3. Equivalents

4.3.1. Equivalents non attestés

Le croisement des données du corpus avec celles des dictionnaires bilingues permet de découvrir que certains équivalents fournis par les dictionnaires bilingues ne semblent pas être en usage. Ainsi, un des deux équivalents que le dictionnaire *OXHA* donne pour *dormant volcano* est le terme *volcan au repos* (l'autre équivalent étant *volcan en sommeil*), tandis que le dictionnaire *HAR* donne, à côté également de *volcan en sommeil, volcan en repos*. Etant donné que ces équivalents, *volcan au repos* et *volcan en repos*, nous paraissaient un peu douteux, nous avons effectué une recherche dans le corpus comparable, et n'avons trouvé aucune attestation de ces termes. En revanche, en faisant une recherche qui combinait les termes *volcan* et *repos*, nous avons trouvé les contextes suivants :

> (11) alternant avec des <u>phases de repos</u> qui peuvent être très longues. Le volcan est alors dit éteint.
> (12) le volcan, qui ne connaît que des <u>phases de repos</u> rarissimes, entre en éruption.
> (13) tenu pour toujours actif un volcan dont l'actuelle <u>période de repos</u> n'est pas considérablement plus étendue
> (14) le volcan napolitain se trouve dans une <u>phase de repos</u> qui n'augure rien de bon pour le futur.

Ces contextes semblent indiquer qu'un équivalent correct serait plutôt *volcan en phase de repos* ou *volcan en période de repos*, mais celui-ci n'est pas attesté tel quel en corpus. Le plus frappant reste toutefois qu'aucun des dictionnaires ne donne *volcan endormi*, qui apparaît 11 fois dans la partie française du corpus comparable. Le corpus traduit, quant à lui, ne contient

que 3 occurrences de la collocation *dormant volcano*, laquelle est une fois traduite par *volcan endormi*, une fois la traduction de *volcan en sommeil* et une fois la traduction de *volcan endormi*. Ni *volcan en repos*, ni *volcan au repos* ne figurent dans ce corpus.

4.3.2. Termes dont la traduction est problématique

Certaines catégories de termes sont particulièrement difficiles à traduire, soit pour des raisons strictement linguistiques, soit pour des raisons conceptuelles.

Les termes métaphoriques, comme le verbe *se réveiller*, sont assez problématiques car il est difficile à un utilisateur de savoir si la métaphore se rend à l'identique dans l'autre langue[24]. En ce qui concerne la collocation *volcan / se réveiller*, il n'est pas évident de déterminer quel équivalent choisir. L'*OXHA* propose à juste titre *to become active again,* mais le *HAR* ne propose aucun équivalent sous *se réveiller*. On trouve cependant *the renewed rumblings* ou *the awakening of a volcano* sous le terme *réveil* de ce même dictionnaire. La deuxième traduction peut paraître légèrement douteuse étant donné qu'elle ressemble fort à un calque, et il pourrait s'agir d'un phénomène de « translationese ». Une analyse du corpus traduit révèle que le terme *awakening*, ainsi que les verbes *awake/re-awaken*, sont utilisés pour traduire *se réveiller* dans les traductions allant du français vers l'anglais, comme dans les trois exemples suivants :

(15) [FRA] Dès leur découverte, la question a été posée de savoir si les volcans d'Auvergne pouvaient **se réveiller**.	[ENG] Since their discovery it has commonly been asked whether the volcanoes of the Auvergne could **re-awaken**. [~]

(16) [FRA] La vengeance des dieux est terrible, le volcan **se réveille**. 1 184 personnes périssent dans les nuées ardentes et les coulées de boue.	[ENG] The vengeance of the gods was terrible, and the volcano - which had been dormant for six centuries – **awoke**: 1184 people died in the ash and mud. [~]

(17) [FRA] Hekla est l'un des volcans les plus actifs d'Islande. Depuis son paroxysme de 1104, qui fut, avec celui de l'Oraefajokull en 1362, la plus forte éruption explosive des temps historiques dans ce pays, il a connu 167 **réveils**, le dernier en janvier 1991.	[ENG] Hekla (above, in a 16th-century print) is one of Iceland's most active volcanoes. Since its eruption in 1104 it has **awakened** 167 times, most recently in January 1991. [~]

Toutefois, le verbe *reawaken* ou le gérondif *reawakening* sont également utilisés dans les textes anglais qui servent de textes sources allant de l'anglais vers le français, comme dans les exemples suivants :

(18) [ENG] The rapid scientific response to the **reawakening** of Mount St. Helens in 1980 drew on earthquake and ground movement-monitoring techniques developed at Kilauea. Similarly, techniques developed at Kilauea – especially the study of earthquake	[FRA] Les techniques de surveillance des séismes et des mouvements du sol mises au point au Kilauea ont permis aux géologues de réagir rapidement face au **réveil** du mont Saint-Helens en 1980, et de prévoir la gigantesque éruption du mont Pinatubo. en

[24] Ce point est également analysé dans Frérot & Josselin [2008].

patterns – were critical to the successful prediction of the climactic eruption of Mount Pinatubo in June 1991, the second or third most powerful volcanic event of this century.	1991 – la deuxième ou la troisième éruption volcanique de ce siècle par sa puissance. [~]

(19) [ENG] Thus any small-scale activity that might have preceded its **reawakening** went unnoticed.	[FRA] En conséquence, toute marque d'activité même faible, qui avait peut-être précédé son **réveil**, passa inaperçue. [~]

L'analyse des données du corpus comparable anglais confirme également que *reawaken* est utilisé par des locuteurs natifs. De plus, le parcours des contextes permet d'enrichir les équivalents possibles. En effet, le parcours du corpus comparable anglais nous a permis de repérer un grand nombre d'équivalents potentiels qui peuvent compléter la miscrostructure du dictionnaire bilingue : *to be again in eruption, to be about to erupt, to be about to blow, to rumble back to life, to threaten to erupt, to begin to erupt, to erupt again, to erupt into life*, etc.

4.3.3. Equivalents non fournis

Comme nous l'avons vu plus haut, le verbe *erupt* ne peut pas toujours se traduire par *entrer en éruption*. L'équivalent semble dépendre de la nature du sujet du verbe, que celui-ci soit transitif ou intransitif. Notamment, un des contextes problématiques anglais était le suivant :

(20) In July 1986, just as it appeared that Puu Oo would start the 48th episode, magma began to erupt from new fissures at the base of the cone.

Une recherche dans la partie française du corpus comparable effectuée en cherchant des contextes associant *magma* ou *lave* et *fissure* a produit les résultats suivants :

(21) Du magma s'épanche de cette fissure ; (22) de la lave rouge-orange sort d'une fissure ; (23) les coulées proviennent d'une fissure

Les trois verbes *s'épancher, sortir, provenir* paraissent donc être des équivalents possibles de *erupt* intransitif dans le cas où celui-ci a pour sujets *magma*, *lava*, ou *flow* et est suivi de la préposition *from*.

Quant au verbe transitif, le corpus comparable nous a livré les équivalents potentiels suivants :

(a) **X erupts Y**
(24) La fissure émet des laves
(b) **Y is erupted (by X)**
(25) Au total, 2 km3 de lave ont été émis
(c) **Y erupted (from X / by X)**
(26) La lave issue de la fissure, (27) lave vomie par la fissure, (28) laves très fluides se déversant par des fissures de l'écorce

Nous résumons les données fournies par le corpus traduit pour le verbe intransitif[25] dans le tableau qui suit :

[25] Rappelons qu'il n'y a qu'une occurrence du verbe transitif dans le corpus traduit. Le verbe *erupt* y est alors traduit par *cracher*.

Equivalent ou procédé de traduction	**Nb occ.**	**Type de sujet**	**Exemple de contexte anglais**	**Exemple de contexte français**
transposition avec *éruption* (*l'éruption du, en éruption)*	17	*volcano, nom de volcan*	Perbuwatan continued to erupt intermittently [...]	Ensuite, les éruptions intermittentes de Perbuwatan [...]
			In 1906 he was actually at Vesuvius when it erupted	Il est en 1906 au chevet du Vésuve en éruption
faire éruption	4	*magma, lava*	Lava from the reservoir erupts to the surface	La lave fait éruption du réservoir jusqu'à la surface
avoir une éruption	1	*volcano*	The volcano had not erupted for 5 000 years	Le volcan n'avait pas eu d'éruption depuis 5 000 ans
entrer en éruption	4	*nom de volcan, volcano*	In the future, Mount St. Helens probably will erupt violently	Dans l'avenir, le Mont Saint-Helens entrera probablement en éruption
[sous-traduction]	4	*ash, lava*	Hardly any fine ash erupts	Les cendres fines sont peu abondantes
s'échapper (du volcan)	2	*magma*	[...] the magma would force its way to the surface and erupt	[...] le magma a forcé son passage et s'est échappé du volcan
sortir	2	*lava*	He even predicted where the lava would erupt	Il annonce où la lave va sortir
s'épancher	1	*basalt*	An alkaline, relatively fluid basalt that erupts at the earth's surface	Un basalte alcalin relativement fluide s'épanchant à la surface de la terre [...]
émise	2	*lava*	Where lava erupts at the surface [...]	[...] la lave émise
rejeter	1	*magma*	Some of the magma may erupt as a tremendous flood of lava	En surface, une quantité considérable de lave serait rejetée
jaillir	4	*magma, lava*	Magma from the secondary reservoir has been erupting along [...]	Le magma issu du réservoir secondaire jaillit le long de [...]
expulser	1	*magma*	When the erupting magma includes little gas	Lorsque le magma expulsé contient des gaz dissous

Tableau 12. Analyse de la construction intransitive du verbe *erupt* dans le corpus traduit

Prenant en compte tous les résultats de cette analyse détaillée du verbe *erupt*, nous avons constitué ce qui pourrait être l'entrée du verbe *erupt* dans un dictionnaire bilingue vraiment basé sur corpus. Cette entrée, que nous avons initialement présentée dans Frérot & Josselin [2008], sépare l'usage intransitif de l'usage transitif, pour lequel elle montre les trois constructions possibles. Elle montre comment les équivalents varient selon les sujets et les objets typiques du verbe, et illustre l'utilisation des équivalents fournis, en proposant parfois d'autres procédés de traduction (comme la transposition) dans les exemples, lesquels sont introduits par l'astérisque simple*. Les collocations verbales sont introduites par **. Cette entrée reflète l'évolution de la lexicographie vers une « lexicographie de l'encodage » telle que la décrit Béjoint [2007 : 46-47], du dictionnaire « Meccano » (c'est-à-dire du « dictionnaire chargé de fournir à ses utilisateurs des éléments lexicaux qui peuvent être alignés bout à bout sur la chaîne syntagmatique pour produire du discours ou pour l'interpréter ») au dictionnaire « patron » (c'est-à-dire « un dictionnaire dont le rôle est de fournir à ses utilisateurs des modèles de discours imitables plus ou moins prêts à l'emploi »).

ERUPT
1) **V.I.** *(volcano)* entrer en éruption * **he observed Vesuvius as it erupted** il observa le Vésuve en éruption ***Hualalai last erupted in 1801** la dernière éruption du Hualalai remonte à 1801 ** **begin to erupt, erupt again** se réveiller; *(magma, lava)* jaillir, s'épancher, sortir * **magma began to erupt from new fissures at the base of the cone** du magma a commencé à s'écouler depuis la base du cône ; *(any other material)* s'échapper, jaillir * **over a four-year period, hundreds of nuée ardentes erupted from Mt. Unzen's summit area** en l'espace de quatre ans, des centaines de nuées ardentes se sont propagées depuis le sommet du Mont Unzen
2) **V.T.** *(magma, lava)* vomir, émettre * **the speed of a lava flow is determined by the volume of lava that is erupted** la vitesse d'une coulée de lave est déterminée par le volume de lave émise **;** *(any other material, particularly ash)* éjecter, émettre, cracher * **Mount St. Helens explosively erupted ash** le Mont St Helens a craché de la cendre en explosant ***the volcano is built up by the accumulation of material erupted through the conduit** le volcan se construit par accumulation des débris éjectés par la cheminée * **Volcanoes on other planets do not necessarily erupt the same materials as Earth's volcanoes** les éjecta des volcans d'autres planètes ne sont pas forcément les mêmes que ceux de la Terre.

Figure 15. Proposition d'entrée de dictionnaire bilingue pour *erupt, v.*

Conclusion

A l'issue de cette analyse croisée des données extraites d'un corpus de vulgarisation et des données tirées des dictionnaires généraux, que peut-on résumer de l'apport spécifique d'un tel corpus dans le processus de rédaction dictionnairique ?

En ce qui concerne la macrostructure, notre expérience révèle que le lexicographe a beaucoup à gagner en comparant les nomenclatures de dictionnaires existants avec des listes de termes extraits de corpus, comme le suggérait d'ailleurs Grundy [1996 : 137]. D'une part, le corpus peut permettre d'identifier certains types de termes qui n'apparaissent pas dans les dictionnaires. Outre le fait qu'il fournit une abondance de collocations que le lexicographe ne pourrait obtenir en se basant uniquement sur la connaissance intuitive qu'il a de sa langue, le

corpus peut ainsi permettre de répertorier avec une plus grande finesse les termes de certaines séries conceptuelles, de répertorier des usages métaphoriques propres à un domaine (pensons à *colère* utilisé clairement dans le sens d'« éruption » en volcanologie), de répertorier des termes empruntés (comme *nuée ardente* en anglais ou *hornito* en français) ou correspondant à une réalité propre à une culture (*puy*, *aa*) ; il permet également d'avoir un aperçu de toutes les constructions dérivées des termes et de balayer l'ensemble des catégories grammaticales. D'autre part, il peut aider le lexicographe à répertorier les variantes des termes, comme les synonymes (*volcan en sommeil* et *volcan endormi* par exemple). Enfin, il peut aussi aider le lexicographe à choisir la forme du mot-vedette, laquelle sera une forme attestée (ce qui amènerait à rejeter *lapillus* par exemple) et la forme la plus courante (ainsi, *volcanologue* à la place de *vulcanologue*). De manière générale, le corpus peut éclairer le lexicographe qui se trouve confronté au choix « entre d'une part des termes dont le rôle est capital dans la compréhension de la structure conceptuelle du domaine et d'autre part des termes dont la plus grande valeur réside dans leur fréquence d'emploi », comme l'explique Thoiron [1998 : 647].

Toutefois, il est indéniable que c'est au niveau de la microstructure que le corpus est d'une plus grande aide par la diversité des contextes qu'il permet de consulter en un temps record. Ainsi, le corpus permet de compléter la description grammaticale des termes, de fournir des définitions quasi-prêtes à utiliser pour les dictionnaires unilingues ou servant de base pour les indications sémantiques des dictionnaires bilingues, de vérifier l'exactitude des équivalents proposés par certains dictionnaires (ex. *volcan au repos*), d'identifier un plus grand nombre d'équivalents (pensons à toutes les traductions possibles de *se réveiller*), de mieux présenter la distinction entre les différents équivalents possibles en utilisant les collocatifs les plus typiques comme indicateurs de contexte (voir le cas d'*erupt*). Plus globalement, le corpus permet d'observer la mise en discours du terme (degré de spécialisation, variation en discours, collocations), dimension essentielle en terminologie comme nous l'avons souligné en introduction.

Pour autant, pour reprendre les mots de Béjoint [2007 : 47], « le corpus est-il la panacée de la lexicographie ? » La réponse est claire : le corpus ne fait pas tout. Malgré tous les efforts mis dans sa constitution, il arrive que certains termes essentiels n'y figurent pas, car il est difficile de réunir un ensemble de textes parfaitement représentatifs du domaine visé. Toutefois, Grundy [1996 : 135] prêche en faveur de l'utilisation en lexicographie générale de sous-corpus spécifiques pour « le vocabulaire hautement technique et spécialisé et le vocabulaire littéraire », expliquant qu'il est « difficile de mettre au point des définitions et des équivalences de traduction pour les termes spécialisés [pour lequel] un corpus de textes de langue générale ne peut pas être très utile ». Il est vrai aussi, comme l'explique Béjoint [*ibid.*], qu'« un corpus coûte cher et [qu']il faut le rentabiliser » ; la question du temps pris par la constitution et l'exploitation du corpus reste en effet posée. Mais si le lexicographe se prémunit d'autant de patience qu'un orpailleur, lorsqu'il découvre en corpus une telle pépite qu'une construction grammaticale répertoriée nulle part (songeons au verbe *erupt*), le jeu n'en vaut-il pas la chandelle ?

Remerciements

Merci à Josette Rebeyrolle et à Anne Przewozny-Desriaux pour leurs précieux conseils et leurs relectures attentives.

Bibliographie

ALTENBERG Bengt and **GRANGER** Sylviane, "Recent Trends in Cross-Linguistic Lexical Studies", in *Lexis in Contrast. Corpus-Based Approaches*, ALTENBERG B. and GRANGER S. (dirs.), Amsterdam, Philadelphie, John Benjamins, 2002: 3-48.

ATKINS Beryl T. Sue, "A Corpus-Based Dictionary", in *The Oxford-Hachette French Dictionary*, CORRÉARD M.-H. and GRUNDY V. (dirs.), Oxford, Paris, Oxford University Press, Hachette, 1994: xix-xxvi.

BEJOINT Henri, "Scientific and Technical Words in General Dictionaries", *International Journal of Lexicography,* 1:4, 1998: 354-368.

BEJOINT Henri, « Nouvelle lexicographie et nouvelles terminologies : convergences et divergences », *Lexicographie et terminologie : compatibilité des modèles et des méthodes,* L'HOMME M.-C. et VANDAELE S. (dirs.), Ottawa, Presses de l'université d'Ottawa, 2007 : 29-78.

BOULANGER Jean-Claude, « L'Aménagement des marques d'usage technolectales dans les dictionnaires généraux bilingues », in *Les Dictionnaires de langue française. Dictionnaires d'apprentissage. Dictionnaires spécialisés de la langue. Dictionnaires de spécialité.* PRUVOST J. (dir.), Paris, Honoré Champion, 2001 : 247-271.

BOULANGER Jean-Claude et **L'HOMME** Marie-Claude, « Les Technolectes dans la pratique dictionnairique générale. Quelques fragments d'une culture », *Meta,* 36:1, 1991 : 23-40.

BOURIGAULT Didier et **CONDAMINES** Anne, « Alternance nom/verbe : explorations en corpus spécialisés », in *Sémantique du lexique verbal, Actes de l'atelier de Caen, 22-23 janvier 1999*, VICTORRI B. et FRANÇOIS J. (dirs.), Caen, Université de Caen, 1999 : 41-48.

BOURIGAULT Didier et **SLODZIAN** Monique, « Pour une terminologie textuelle », *Terminologies nouvelles,* 19, 1999 : 29-32.

BOURIGAULT Didier, **FABRE** Cécile, **FREROT** Cécile, **JACQUES** Marie-Paule et **OZDOWSKA** Sylwia, *Syntex, analyseur syntaxique de corpus*, *Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles*, Dourdan, 2005, publication en ligne.

BOWKER Lynne and **PEARSON** Jennifer, *Working with Specialized Language, A Practical Guide to Using Corpora*, Londres, New York, Routledge, 2002.

CANDEL Danielle, « La Présentation par domaines des emplois scientifiques et techniques dans quelques dictionnaires de langue », *Langue française,* 43, 1979 : 100-118.

CANDEL Danielle (dir.), *Français scientifique et technique et dictionnaire de langue*, Paris, Didier, 1994.

CANDEL Danielle et **HUMBLEY** John, « Explorations terminologiques dans un dictionnaire de langue. Domaine : géologie », in *Mélanges de linguistique offerts à Rostislav Kocourek*, LAPIERRE L., OORE I. et RUNK H. R. (dirs.), Halifax (N. S.), Les presses d'ALFA, 1997 : 35-48.

COTTEZ Henri, « La Place des termes techniques et scientifiques dans un dictionnaire de langue », in *Français scientifique et technique et dictionnaire de langue*, CANDEL D. (dir.), Paris, Didier Erudition, Collection « Etudes de sémantique lexicale », 1994 : 13-26.

DELAVIGNE Valérie, *Les Mots du nucléaire. Contribution socioterminologique à une analyse des discours de vulgarisation*, thèse de doctorat non publiée, Université de Rouen, Rouen, 2001.

FREROT Cécile et **JOSSELIN-LERAY** Amélie, « Contribution des corpus à l'enrichissement des dictionnaires bilingues généraux. Application au domaine de la volcanologie », in *Autour des langues et du langage : perspective interdisciplinaire*, Actes du colloque international des étudiants chercheurs en Didactique des Langues et en Linguistique, 4-

7 juillet 2006, Université Stendhal Grenoble, Presses Universitaires de Grenoble, 2008 : 415-422.

GRUNDY Valerie, « L'utilisation d'un corpus dans la rédaction du dictionnaire bilingue », in *Les dictionnaires bilingues*, BEJOINT H. et THOIRON P. (dirs.), Louvain, Duculot, 1996 : 127-149.

JOSSELIN-LERAY Amélie, *Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues. Etude d'un domaine de spécialité : volcanologie*. Thèse de doctorat non-publiée, Université Lumière Lyon II, 2005.

JOSSELIN Amélie, « Constitution d'un corpus de vulgarisation dans le domaine de la volcanologie : objectifs, méthode et problèmes dans une optique de lexicographie générale », in *Applications et implications en sciences du langage*, LEGLISE I., CANUT E., DESMET I. et GARRIC N. (dirs.), Actes des Journées Jeunes Chercheurs (mars 2003), Paris, l'Harmattan, 2006 : 243-256.

JOSSELIN Amélie et **ROBERTS** Roda P., « La définition des termes dans les dictionnaires généraux unilingues : analyse de quelques exemples du domaine de la volcanologie à la lumière d'un corpus de vulgarisation », in *Lexicographie et terminologie : compatibilité des modèles et des méthodes*, L'HOMME M.-C. et VANDAELE S. (dirs.), Ottawa, Presses de l'Université d'Ottawa, 2007 : 141-188.

LANDAU Sidney I., *Dictionaries. The Art and Craft of Lexicography*, Cambridge, New York, Melbourne, Cambridge University Press, 1989.

LAURIAN Anne-Marie, « Représentation de la science dans les dictionnaires bilingues généraux », in *Le français dans les dictionnaires bilingues,* SZENDE T. (dir.), Paris, Champion – Slatkine, 2006 : 337-346.

L'HOMME Marie-Claude, « Fonctions lexicales pour représenter les relations sémantiques entre termes », *TAL,* 43:1, 2002a : 19-41.

L'HOMME Marie-Claude, *What can Verbs and Adjectives Tell us about Terms?*, Sixth International Conference on Terminology and Knowledge Engineering, Nancy, 28-30 August 2002, 2002b: 65-70.

L'HOMME Marie-Claude, *La Terminologie : principes et techniques*, Montréal, Presses de l'Université de Montréal, 2004.

L'HOMME Marie-Claude et **VANDAELE** Sylvie (dirs), *Lexicographie et terminologie : compatibilité des modèles et des méthodes,* Ottawa, Presses de l'université d'Ottawa, 2007.

MAZIERE Francine, « Le Dictionnaire et les termes », *Cahiers de Lexicologie,* 2:39, 1981 : 79-104.

MEYER Ingrid and **MACKINTOSH** Kristen, "The Corpus from a Terminographer's Viewpoint", *International Journal of Corpus Linguistics,* 1:2, 1996: 257-285.

MORTUREUX Marie-Françoise, « L'Analyse du discours de la vulgarisation scientifique et le dictionnaire de la langue scientifique », in *Français scientifique et technique et dictionnaire de langue*, CANDEL D. (dir.), Paris, Didier, 1994 : 63-75.

MORTUREUX Marie-Françoise, « Les Vocabulaires scientifiques et techniques », *Les Carnets du CEDISCOR,* 3, 1995 : 13-25.

QUEMADA Bernard, *Les Préfaces du Dictionnaire de l'Académie française 1694-1992*, Paris, Honoré Champion, 1997.

REY Alain, « La Terminologie dans un dictionnaire général de la langue française : *Le Grand Robert* », *TermNet News,* 14, 1985 : 5-7.

REY Alain, *De l'artisanat des dictionnaires à une science du mot. Images et modèles*, Paris, Armand Colin, 2008.

RAIMBAULT Jean-Claude, *Si mon dico m'était conté*, Paris, Mots et Cie, 2003.

ROBERTS Roda P., "Methods of Bilingual Dictionary-Making: the Canadian Experience", in *Symposium on Lexicography V. Proceedings of the Fifth International Symposium on Lexicography*, HYLDGAARD-JENSEN K. et ZETTERSTEN A. (dirs.), May 3-5, 1990, University of Copenhagen, Copenhague, Niemeyer, 1992: 91-115.

RODRIGUEZ REINA Maria Pilar, « Les Secteurs de la terminologie maritime dans les dictionnaires italien-espagnol », in *Terminologie maritime : traduire et communiquer*, D. L. Newman & M. Van Campenhoudt (dirs.), Actes du 1er colloque international de terminologie maritime, 15-16 mai 1998, Bruxelles, Editions du Hazard, 1999 : 295-310.

SINCLAIR John, "Lexicographic Evidence", in *Dictionaries, Lexicography and Language Learning*, ILSON Robert (dir.), Oxford, Pergamon Press & The British Council, 1985: 81-92.

SINCLAIR John, *Looking-Up – An Account of the Cobuild Project in Lexical Computing*, Londres, HarperCollins, 1987.

SVENSÉN Bo, *Practical Lexicography. Principles and Methods of Dictionary Making*, Oxford, New York, Oxford University Press, 1993.

TEUBERT Wolfgang, "Corpus linguistics and lexicography", *International Journal of Corpus Linguistics*, 6:1-12, 2001: 125-153.

THOIRON Philippe, « Place et rôle de la terminologie dans les dictionnaires bilingues non-spécialisés. Le cas de la terminologie médicale dans le dictionnaire Hachette-Oxford (français-anglais) », in *Mots chiffrés et déchiffrés*, MELLET S. et VUILLAUME M. (dirs.), Paris, Genève, Honoré Champion, Slatkine, 1998 : 623-650.

WESEMAËL François et **WESEMAËL** Roland, « L'Expression de la science dans un dictionnaire culturel », in *Les dictionnaires Le Robert. Genèse et évolution*, CORMIER M.-C., FRANCOEUR A. et BOULANGER J.-C. (dirs.), Montréal, Presses de l'Université de Montréal, 2003.

Dictionnaires

Dictionnaire de la tectonique des plaques et de la géodynamique, VILA Jean-Marie, Paris, Gordon and Breach Science Publishers, 2000.

Grand dictionnaire encyclopédique en dix volumes, GILLON Etienne, HOLLIER-LAROUSSE Jacques *et al.* (dir.), Paris, Larousse, 1960-1964.

Harrap's Shorter, Paris, Havas Interactive Europe & Harrap's Multimedia, 2000. [CD-ROM]

Le Grand Robert, REY Alain (dir.), Paris, Dictionnaires Le Robert, 1985.

Le Nouveau Petit Robert : Dictionnaire de la langue française, REY-DEBOVE Josette et REY Alain (dirs), Paris, Dictionnaires Le Robert, 2001. [CD-ROM]

Le Petit Larousse Illustré, EVENO Bertrand, LEGRAIN Michel *et al.* (dirs), Paris, Larousse, 2002. [CD-ROM]

Le Petit Larousse Illustré, LEGRAIN Michel, GARNIER Yves *et al.* (dirs), Paris, Larousse, 2003.

The American Heritage College Dictionary, COSTELLO Robert B. (dir.), Boston, New York, Houghton Mifflin, 1996. [CD-ROM]

The Collins COBUILD English Dictionary, SINCLAIR, John (dir.), London, HarperCollins Publishers, 1987.

The New Oxford Dictionary of English on CD-ROM, Oxford, Oxford University Press, 2000.

The Oxford-Hachette French Dictionary / Le Dictionnaire Hachette-Oxford, CORRÉARD Marie-Hélène et GRUNDY Valerie (dirs), Oxford, New York, Toronto, Hachette, Oxford University Press, 1994.

The Oxford-Hachette French Dictionary / Le Dictionnaire Hachette-Oxford, CORRÉARD Marie-Hélène et GRUNDY Valerie (dirs), Oxford, New York, Toronto, Hachette, Oxford

University Press, 1996. [CD-ROM]

The Oxford Dictionary of Earth Sciences (2003). ALLABY Ailsa & Michael ALLABY, Oxford, Oxford University Press.

Trésor de la langue française : Dictionnaire de la langue du XIXè et du XXè siècle (1789-1960), IMBS Paul (dir.), Paris, Editions du Centre National de la Recherche Scientifique, 1971.

Webster's Third New International Dictionary of the English Language Unabridged, GOVE Philip B. (dir.), Springfield (Mass.), Merriam-Webster, 1986.